DOCUMENT RESUME

ED 177 483 CS 005 014

AUTHOR Crothers, Edward J.
TITLE Text Design for Coherence, Comprehension, and Memory. Final Progress Report.
INSTITUTION Colorado Univ., Boulder.
SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C.
PUB DATE 31 Jul 79
GRANT NIE-G-77-0011
NOTE 86p.

EDRS PRICE MF01/PC04 Plus Postage.
DESCRIPTORS *Cognitive Processes; *Discourse Analysis; Higher Education; Memory; *Reading Comprehension; *Reading Research

ABSTRACT

A series of experimental and theoretical investigations of text inference were conducted. One applied a theory of inferable text cohesion to the devising of text structure revisions that assist comprehension and memory. Since only modest success was achieved, the next three projects went beyond text revision to the use of adjunct outlines, questions, and connectives in novel formats. The last of these projects yielded the most pronounced effects on responses and response times. Methods, materials, and computer programs for testing comprehension during and after the reading were developed. Subjective text cohesion reflected by performance on a sentence sequencing task was compared against theoretical representations. A grasp of causal and logical connectives within a text was found to depend on the interaction of the relation's directness and direction. Advances in text representation theory also were accomplished. After a digression on formalizing a text as a "proof," a deeper description highlighting parallelism was conceived. (Author/TJ)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

ED177483

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

FINAL PROGRESS REPORT

GRANT NO. NIE-G-77-0011

# TEXT DESIGN FOR COHERENCE, COMPREHENSION, AND MEMORY

EDWARD J. CROTHERS
UNIVERSITY OF COLORADO
BOULDER, COLORADO 80309

JULY 31, 1979

The research reported herein was performed pursuant to a grant with the National Institute of Education, U.S. Department of Health, Education and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official National Institute of Education position or policy.

U. S. DEPARTMENT OF

HEALTH, EDUCATION AND WELFARE

NATIONAL INSTITUTE OF EDUCATION

TEACHING AND LEARNING

CS00501K

## Abstract

The research consisted of a series of five experimental and three theoretical investigations of text inference.

One experimental project applied the author's theory of inferable text cohesion to the devising of text structure revisions that assist comprehension and memory. Only modest empirical success was achieved, however. Therefore the next three projects went beyond text revision itself, to the use of adjunct outlines, questions, and connectives in novel formats. Of these projects, the last yielded by far the most pronounced effects on responses and response times. The remaining experimental project turned from educational applications to psychological processes. Methods, materials, and computer programs for testing comprehension during and after reading were developed. Subjective text cohesion reflected by performance on a sentence sequencing task was compared against theoretical representations. Also, grasp of causal and logical connectives within a text was found to depend on the interaction of the relation's directness and direction.

Advances in text representation theory also were accomplished. After a digression on formalizing a text as a "proof", a deeper description highlighting parallelism was conceived. Presently a more abstract, powerful formulation of this description is being developed.

## Preface

This report is organized as a series of eight projects, each consisting of one or more experimental or theoretical investigations. Methods and findings are reported individually for each of these projects; for the experimental projects all results, discussion, and conclusions appear under the heading Results.

Publications to date have presented the evolution of the theoretical side of the work (Crothers, 1978, 1979a, 1979b). When completed, the current theoretical advance (the last project reported herein) will also be submitted for publication, as will the most promising experimental findings (the last two experimental projects).

In future continuation of the research beyond the grand period, top priority will be given to the final sentence sequencing and theory development projects reported.

Table of Contents


Methods and Findings 1

I. Experimental comparisons of memory for original and revised texts 1

Experiment 1 1
Experiment 2 4
Experiment 3 6
Experiment 4 7
Experiment 5 8

II. Using thematic information and text revisions to improve the effectiveness of adjunct questions 14
Experiment 6 14

III. Using adjunct outlines in an attempt to assist text memory 16
Experiment 7 16

IV. Using adjunct connections to assist text comprehension and memory 22
Experiment 8 22

V. Text processing issues examined using the Circle Island narrative 34

Experiment 9 34
Experiment 10 38
Experiment 11 40
Experiment 12 42
Experiment 13 55
Experiment 14 57

VI. Toward representing a text as a "proof" in natural language logic 58

VII. Representing parallelism in texts: Circle Island 67

VIII. Representing text interpretations by means of list graphs 75

References 77

## Methods and Findings

I. Experimental comparisons of memory for original and revised texts (Edward Crothers, assisted by Mark Fetler and David Williams).

The major goal was to apply the author's descriptive theory of text structure (Crothers, 1978, 1979, a,b) to the developing of principles for revising texts' structures so as to enhance the memorability (and by the same token comprehensibility in the first place) of a given content. Because these principles involve text structure only, rather than other aspects (e.g. adjuncts such as tables, figures, summaries, or outlines), the revisions in turn were limited to the structure. Five studies, counting the pilot ones, were completed under the project. Highlights of the methods and findings are summarized very briefly in Table 1. The last study in the series, Experiment 5, culminated the work and yielded most of the major conclusions.

Experiment 1.

An initial expectation was that the benefit from structure revisions would of course depend on the nature of the original text prior to revision. At one extreme, a text already optimally organized leaves little room for further improvement. At another extreme, if a text is both complex and technical (as judged by a mature reader who however is not a specialist in the particular subject matter), then the revisions required to clarify it would presumably go well beyond purely structural changes that are the realm of the present theory. In such a case one would need an entire theory of instruction in order to specify the prerequisite subject matter that must be taught to prepare the reader for the technical passage. But intermediate between these two extremes is the presumably frequent case where text structure revision would by itself suffice to promote memory. Therefore, although the hypothesis was that an effect of text version (Revised vs. Original) would be found generally over different texts, it was also anticipated that Version x Text interactions would emerge.

Method. Two popularly styled scientific passages were selected for this and later studies. One was part of a Scientific American article on

Table 1

Summary of studies using conventional text format and Version (O,R) x Text (N,C) x Order (1,2) design[a]

| | Exp. 1(3/77) | Exp. 2(3/77) | Exp. 3(6/77) | Exp. 4(9/77) | Exp. 5(12/77) |
|---|---|---|---|---|---|
| **Method** | | | | | |
| a. Instructions | standard | to expect difficult Test | as in Exp. 2 | as in Exp. 2 | as in Exp. 2 |
| b. Training | $(SST)_1$, $(SST)_2$ | $(SST)_1$, $(SST)_2^1$, $(SST)_2^1$, $(SST)_2^1$ | as in Exp. 1 | as in Exp. 1 | as in Exp. 1 |
| c. Pacing | self- | experimenter- | as in Exp. 2 | as in Exp. 2 | as in Exp. 2 |
| d. Revisions (of structure only) | multiple types, moderate extent | same | multiple types, more sweeping | as in Exp. 3 for C text | as in Exp. 4 |
| e. Tests | 40-item fill-in 12-item mult. choice | slightly refined | further refined | further refined | further refined |
| f. Special features | -- | session 24 hrs. apart | -- | revisions from "proof" theory for N text | as in Exp. 4 |
| **Data** | | | | | |
| a. overall level of performance | low | high | moderate | moderate | moderate |
| b. Sig. effects re version | R2>02 | RC>OC | none | RN>ON | O>R |
| c. Other | effect only on fill-in test | as in Exp. 1 | -- | as in Exp. 1 | No effect on N fill-in |

[a]Abbreviations: O = original version, R = revised version, N = Nobel Prize text, C = Catastrophe Theory text, R2 = revised text presented second, etc.; S = study, T = test.

1976 Nobel prize winners. Basically, the article is a summary of the significance of their work, not a technical exposition of it. The other passage was the first part of a somewhat more abstract and technical Scientific American article, this one on mathematical catastrophe theory. By excerpting only the first part it was possible to avoid the deeper complications of the exposition, concentrating instead on an informal introduction to its purpose and main notions. Both passages were about 500 words long.

Revised versions of both passages were then prepared. Although the theory aids in identifying potential revisions, only a mixture of intuition and experimentation can determine which revisions will in fact aid comprehension. The hypothesis, at least, was that the most effective revisions would be those that contribute in linguistically describable ways to the coherence of the text. However, in initial pilot work it seemed premature to thus restrict the scope of revisions. Therefore other types, such as a few definitions of unfamiliar terms, were also included. Each revision was described in detail by the theory, both to specify the revision precisely and to allow for later investigation of which kinds of revisions might prove most effective. Besides preparing the two revised texts, the other major task was the preparation of the memory tests. Two tests, one a fill-in and the other a multiple-choice, were constructed for each passage. Without going into full detail here, the tests can be generally described as follows. Each fill-in test contained about 40 items, each multiple choice test about 12. The fill-in tests emphasized competence in reproducing or paraphrasing the original text, in both its main ideas and its details. By contrast, some questions in the multiple choice test were phrased to test for simple inferences from the texts, though neither text was particularly notable for significant inferences. An important aspect of the tests was the correspondence between each test item and the critical part of the passage. There are two possibilities. Either the item derives from a part of the text that has undergone one or more revisions, or it comes from a part where the revised version is identical to the original. Hence one can of course compare test data from the original and revised versions on an item-by-item basis, instead of merely comparing them on total test scores.

College students from introductory psychology served as experimental participants. They were run one or two at a time. Each subject served in one cell of a Latin square design: the factors were passage (Nobel Prize or Catastrophe Theory), order of passage (First or Second), and version (Original or Revised). The constraints on the design were that each subject read one original and one revised passage. For example, a subject in one cell of the Latin square might receive the revised Nobel passage first, then the original Catastrophe passage. The study used eight subjects, two randomly assigned to each cell of the Latin square. The sequence of events was: reading instructions, read first passage twice, first fill-in test, first multiple choice test, then repeating with the second passage. Besides providing some preliminary memory data, the other objective of this first study was to estimate reading times per passage. Reading time data are essential, because most of the revisions tend to lengthen the passage. Therefore in later studies, to control for speed-accuracy tradeoffs, it was necessary to allow all subjects the same total reading time. The present data allowed us to set the value of this time; eight minutes for the two readings of a passage proved to be the least duration sufficient for all but the slowest reader.

Results. The findings were intended merely as a clue to guide future research, and hence were subjected only to overall data analyses. The effect of version was slight and, for practical purposes, nonexistent. On the fill-in tests the revised version yielded only slightly higher overall performances than did the original. Primarily, the superiority of the revised version was limited to the case where it was the subject's second passage. On the multiple choice tests there was no overall effect of revision. Actually, the most striking feature of the data was the overall low level of performance, averaging no more than 40% on the fill-in test and 60% on the multiple choice test. It was conjectured that a "floor" effect may have contributed to the negligibility of the version effect.

Experiment 2.

Now the objective was to bring the readers to a higher level of mastery, in order to see if a version effect would then emerge.

Method. Several changes from Experiment 1 were introduced: (i) A sentence was inserted in the

instructions specifically warning subjects to expect exceptionally comprehensive memory tests. (ii) Now all stages of the procedure were experimenter-paced, in order to ensure equal reading times for both original and revised versions. As planned, the durations were set at values which, judging from the times in the first study, would be adequate for nearly all subjects to complete each stage. To accommodate the faster readers, subjects who finished early were instructed to go back and review the text. (Presumably more self-review occurred with the shorter, original versions than with the revised versions.) Thus, this particular way of implementing the experimenter-pacing allowed most readers more time than they would have had with self-pacing, hence was expected to yield higher test scores. (iii) A major procedural change was that, after the test on a passage, a second reading-testing cycle identical to the first was now introduced. Its purpose was to examine the idea that perhaps one has to be tested on a passage first in order to become sensitive to the organizational clarifications in the revised version. If so, the superiority of revised over original should be more marked on the second test cycle than on the first. (There is an obvious counter-hypothesis, however. Repeated reading and testing might have exactly the opposite outcome, for two reasons: It might produce almost maximal performance under both text versions. Also, administering an identical test to subjects who had received different reading versions would amount to reducing the total difference between the materials read by the various subjects, thus possibly leading to a convergence in performance on the retest.) This third change almost doubled the duration of the experiment. Therefore the session was split into two one-hour sessions, 24 hours apart. All reading and testing of the subject's first or second passage occurred in the first or second session, respectively. (iv) Finally, minor modifications were made in the four sets of test questions. Data from the first study indicated that a small percentage of the questions were either ambiguous or else too difficult. Hence such questions were reworded or deleted. Other than on points (i) - (iv) above, the method was identical to the first pilot experiment. In particular, the same Latin square design was used, randomly assigning two subjects to each of the four cells of the square.

Results. Findings were mixed. On Catastrophe Theory the revised version did surpass the original, and an item analysis showed that the effect extended to most test items. However, on Nobel Prize no effect of version was found. Another negative outcome was that

the retesting produced a ceiling effect instead of the hypothesized interaction with version.

At this point it was decided to reconsider the previously dismissed possibility that perhaps the negative results were due not so much to the notion of structure revisions per se but to some overly conservative practice in composing the revisions. In other words, the theory only identifies inferences, the explication of some of which might be expected to aid comprehension. But there are different ways and degrees that one can implement the explications, and on this educational strategy the theory is mute.

Experiment 3.

Method. This time the structure revisions were composed in forms that altered the original text more radically. More explications and recapitulations were inserted, for the first time substantially lengthening the text compared to the original version. The design was as before. Six participants served in each of the usual four conditions, namely, (retaining the Table 1 notation) NR-CO, NO-CR, CR-NO, and CO-NR. The procedure required about 40 minutes. Other details of it are summarized in the table. Test questions again followed the "shotgun" approach of testing all kinds of information: both explicit and implicit, both details and themes, as well as both revised and unrevised segments of the text.

Results. Surprisingly, results were uniformly negative with respect to the version variable. No difference was manifest on either test or either passage.

Possible conclusions at this point:

i. Coherence-enhancing revisions are helpful only if the original text is severely noncoherent.

ii. As some would argue, comprehension depends far more on STM and/or LTM than on text structure.

iii. The "shotgun" strategy of making diverse revisions and testing all is, despite its educational practicality, an insensitive approach for detecting version effects on memory.

Conclusions (i) and (ii) would amount to retreating from the initial hypothesis, while conclusion (iii) would limit its practical significance. Hence before embracing any of these conclusions it was decided to undertake a critical reappraisal of the

theoretical analysis itself. This detour was also independently motivated by purely theoretical reasons. The new theoretical direction that sprung from these two concerns is summarized elsewhere herein (parts VI and VII); the ensuing experiment will now be described.

Experiment 4.

This was a preliminary exploration of the possibility that revisions suggested by the new theoretical venture (dubbed a "proof" representation of texts) might aid memory.

Method. As an expedient, because the "proofs" are time-consuming and somewhat intricate to write out and also because a version effect had already been demonstrated for Catastrophe Theory, the new method was applied only to generate the Nobel Prize revisions. As shown in Table 1, the method otherwise followed that of the preceding experiment. Twelve college student volunteers served as participants, three in each of the four cells of the design.

As for the new Nobel Prize text revisions, they were indeed nontrivially different from those motivated by the original theory. To give a general idea, main conclusions were placed before their proofs, something that only occurred in a hit-or-miss fashion in both the original and earlier revised versions. Apart from this, a noteworthy change was that the new version was trimmed of excess obvious inferences that had been inserted in the previous revision for the purpose of explicating the coherence. The effect was to return the length of the revision back again to virtually that of the original text prior to any revision. This sort of trial-and-error adjustment of the explicitness of the text is another example of the fact, already emphasized, that a theory of structure is not ipso facto a theory of optimal presentation of that structure.

Results. The outcome was more encouraging this time. The new revisions, those of the Nobel Prize text, led to 82% correct on the 40-item short-answer memory test compared to only 62% for participants receiving the original version. The effect, however, was slight on the less comprehensive, more inferentially oriented multiple-choice test; percentages correct were 87% and 81%, respectively. At the same time, the old revisions, those of Catastrophe Theory, yielded results consistent with the previous study; there was no improvement from original to revised version. The

positive results with Nobel Prize led us to follow up this pilot study with a full-fledged experiment, next.

Experiment 5.

A larger scale replication of the preceding experiment was now conducted.

Method. The method adhered to that summarized above. Some particular details worth noting will now be added to the description.

Twenty participants were assigned randomly to each of the four cells of the design. The Nobel original and revised versions were about 1½ and 2 double spaced typed pages in length, respectively. Corresponding lengths for the Catastrophe text were 2¼ and 2½ pages, respectively.

Each participant also received four test booklets, namely a fill-in-the-blank test and a multiple choice test on each passage. The fill-in tests were intended to be fairly comprehensive tests of the explicit content. Both thematic and detailed questions were included. The exact wording of the fill-in questions was somewhat of a problem, because in practice it is difficult to word each question in a way that is equally near the two versions, original and revised, for which memory is being tested. Therefore a conservative expedient was adopted, namely wording the fill-in questions generally conforming to the corresponding wording of the original, rather than revised, version. The sequence of questions also adhered to the sequence of corresponding answers in the original version. The particular words questioned in each fill-in item were the main words of new information introduced in the corresponding text sentence. The pool of questions had been refined over the course of several previous studies, thereby eliminating questions that proved to be ambiguous. Because the hypothesis was that the memory improvement would extend to all levels of text statements, questions on all levels were included. The Nobel fill in-test consisted of 40 blanks distributed among 29 questions; for Catastrophe Theory the corresponding frequencies were 33 and 22.

The multiple choice test was intended to offer questions whose wording differed greatly from the surface text. Originally, a further intention was to test only higher level information, not only explicit but implicit content as well. However, only a limited number of such items could feasibly be constructed,

owing in part to the difficulty of conceiving three plausible but incorrect alternatives per question. Therefore a diversity of multiple choice items including both theme and detail and both explicit and implicit content were included. There were 14 questions for Nobel and 12 for Catastrophe.

Experimental events proceeded in the following sequence: initial reading instructions, reading of first passage, re-reading of first passage, fill-in test instructions and test on first passage, multiple choice instructions and test on first passage, then repeating with the second passage. In the initial reading instructions, subjects were told of the stages to follow, specifically including the facts that there would be two comprehensive tests after each passage. Under each condition subjects were given seven minutes to read and reread a passage. Subjects were told to use the full seven minutes, by further reviewing the text if they finished rereading it before the seven minutes elapsed. Each fill-in test occupied eleven minutes and each multiple choice test five minutes.

Results. The hypothesized main effect of version was supported. Percentagewise, the differences between versions were not strikingly great, however. Three of the four tests, namely excluding the Nobel Prize fill-in test, showed higher proportions correct for revised than for original versions (see Table 2). As shown, the overall percentage superiority of revised over original versions was .08 on multiple-choice tests but only .02 on fill-in (short answer) tests. It might also be added that the reversal on the Nobel Prize fill-in test was not clarified when the interaction with presentation order (first or second) was examined; the outcome was about the same under both orders. A more promising lead on this reversal was revealed by a supplementary analysis of test items individually. In particular, of the 40 items on the Nobel Prize fill-in test it turned out that a certain 3 accounted for no less than 73% of the frequency difference in favor of the original version.

Recall that a key feature of the design was that, to reduce variability, the same participant was run with both values of the version variable. This design is known as a fractional factorial replication (Hicks, 1973, Ch. 15). The analysis (performed by David Williams) was as follows. First, the frequencies correct were converted to percentages, to be used in an ANOVA.

Table 2

Proportions Correct in Experiment 5

Short answer tests (N=660 per cell for Catastrophe Theory, 800 for Nobel Prize):

| Group | Catastrophe Theory | Nobel Prize |
|---|---|---|
| Nobel Orig., Catas. Rev. | .63 | .64 |
| Nobel Rev., Catas. Orig. | .46 | .56 |
| Catas. Orig., Nobel Rev. | .48 | .64 |
| Catas. Rev., Nobel Orig. | .57 | .69 |

Total: Catas. Orig. = .47, Catas. Rev. = .60, Nobel Orig. = .67, Nobel Rev. = .60; Orig. = .58, Rev. = .60.

Multiple-choice tests (N=240 per cell for Catastrophe Theory, 280 for Nobel Prize)

| Group | Catastrophe Theory | Nobel Prize |
|---|---|---|
| Nobel Orig., Catas. Rev. | .49 | .72 |
| Nobel Rev., Catas. Orig. | .40 | .78 |
| Catas Orig., Nobel Rev. | .40 | .84 |
| Catas. Rev., Nobel Orig. | .49 | .75 |

Total: Catas. Orig. = .40, Catas. Rev. = .49, Nobel Orig. = .73, Nobel Rev. = .80; Orig. = .58, Rev. = .66

Three factors were of principal interest: Text (Tx), Catastrophe vs. Nobel; Version (Vr), Original vs. Revised; and Test (Ts), multiple-choice vs. fill-in. Other factors were intended primarily to control for extraneous variation: Version order (Vo), receiving the one version before the other; and Time (Tm), ordinally construed as the first or second half of the experimental session. Version, Text, Test, and Time were within-subjects factors, while Version Order and Text Order were between-subjects factors.

The design was thus a ¼ replicate of a six-factor design having the defining relations TmTxTo, TmVrVo, and VrTxToVo. In a fractionally replicated design some effects are confounded with other effects. Therefore any interpretation is justified only if certain effects can reasonably be assumed to be negligible. This is necessary in order for a significant F-value to be attributable to one effect rather than to the sum of several, perhaps individually nonsignificant, effects. In an ideal fractionally replicated design, main effects and those interactions of greatest interest are confounded with interactions of the highest possible order. Because the set of defining relations for the present analysis do not give one of the most efficient fractionally replicated designs (i.e., some main effects were confounded with first order interactions, other first order interactions were confounded with still other first order interactions, and second order interactions were confounded with second order interactions), two points should be made explicit. First, this set of relations was dictated by the nature of the factors: Two factors were the order (first or second) of two other variables, and this order was intentionally counterbalanced. Second, the interpretation is nonetheless relatively unambiguous, because for the effects that turned out significant the confounded interactions can reasonably be assumed negligible.

The outcome of the ANOVA appears in Table 3, the directions of effects having already been indicated in Table 2. The version (Vr) effect was highly significant and in the hypothesized direction. The version variable appeared in only one other significant effect, namely the TxTsVr (text x test x version) interaction. From Table 2 it can be seen that the TxTs interaction was greater for version R than for version O. No other interactions involving Vr reached significance, not even the TxVr interaction. Effects not involving Vr, though of lesser interest here, are worth noting. Not surprisingly, the Catastrophe Theory

Table 3

Results of ANOVA for Fractional
Replication Factorial Design of Experiment 5

| Source | SS | df | MS | F |
|---|---|---|---|---|
| Between Ss | | | | |
| To | 426 | 1 | 426 | .5 |
| Vo | 194 | 1 | 194 | .2 |
| TxVr | 2639 | 1 | 2639 | 2.9 |
| error | 68744 | 88 | 905 | |
| Within Ss | | | | |
| Tx | 36830 | 1 | 36830 | 223.2*** |
| Ts | 503 | 1 | 503 | 3.0 |
| Vr | 2492 | 1 | 2492 | 15.1*** |
| Tm | 928 | 1 | 928 | 5.6* |
| TsVr | 483 | 1 | 483 | 2.9 |
| TsVo | 60 | 1 | 60 | .4 |
| TsTo | 0 | 1 | 0 | .0 |
| TsTx | 10638 | 1 | 10638 | 64.5*** |
| TmTs | 33 | 1 | 33 | .2 |
| VrTo | 6 | 1 | 6 | .0 |
| TxTsVo | 77 | 1 | 77 | .5 |
| TxTsVr | 1697 | 1 | 1697 | 10.3** |
| residual | 37634 | 264 | 165 | |
| Total | 194330 | 367 | | |

* .01 <p<.025
** <p<.005
*** p<.001

passage was more difficult than the Nobel Prize one. Why this effect was greater on the multiple-choice test than on the short-answer test is not clear. A likely conjecture arises from the observation that the short answer tests emphasized names and other more concrete facts, while the multiple choice tended to emphase principles. Names should be relatively easy to remember in any case, but principles should be harder the more abstract the passage is.

In sum, the version hypothesis was confirmed, though not dramatically so. On further thought, however, the outcome is somewhat more impressive. That is, it will be recalled that the tests covered the passage content rather thoroughly, instead of being confined to items that had received more emphasis in the revised version. In fact, some had been phrased identically in both versions and even a few had been subordinated, downplayed, in the revised versions.

Finally, a few reflections are in order as to the method adopted in this series of studies, in particular the blanket or "shotgun" approach to both revising and testing. An obvious advantage of this method is that it facilitates a rapid initial probe of the hypothesis. On the other hand, the method is too crude to allow testing of more refined hypotheses. Not only is there variation between test items in the textual role of the information queried, there is also variation as to whether and how presentation of that information was revised. Additional sources of variation enter when one replicates with a second passage. For natural prose it is difficult to obtain two texts whose structures parallel each other. Hence it can hardly be claimed that the two sets of revisions were at all parallel, or that the two sets of test items were. The research strategy actually adopted followed the philosophy of seeking large gross effects in order to motivate more refined, time-consuming follow-up studies. Unfortunately, the results do not show large effects, even though technically they do support the hypothesis. An opposite research strategy would have followed the philosophy that the refined designs are what bring out the impressive effects. Against this view, I doubt its appropriateness either for the present question or for educational psychology research generally. As to the former, because no detailed hypothesis suggested itself as to exactly which textual types of revisions should be most effective, it seemed reasonable to test many kinds simultaneously. For educational research generally, it can be argued that, in the first place, gross effects are the only ones that have a chance of

holding up in the classroom and that, in the second place, what should be sought is improvement in total performance rather than only in performance on selected test items.

II. Using thematic information and text revisions to improve the effectiveness of adjunct questions (Mark Fetler's dissertation, supervised by Edward Crothers).

Experiment 6.

When not only revisions but adjunct questions as well are used with the aim of improving comprehension, one could hypothesize either that the effects of the two manipulations will be cumulative, or else that the one will in practice obviate the other. Testing of these hypotheses was one goal of this dissertation. The main purpose, however, was the development and testing of a process model on the effects that adjunct questions have on text comprehension and memory. In particular, the model should account for any effects attributable to varying the thematicity of the adjunct question.

Positive effects of adjunct questions on memory for prose are well documented by Anderson and Biddle (1975) and Rothkopf (1972, 1976). A variable as yet relatively poorly understood is the importance or thematicity of the material questioned. Does it make a difference whether the material questioned is thematic, or can one expect equally good improvements in comprehension and memory if relatively unimportant details are questioned? Studies in Kintsch (1977) have shown the effects of the level or degree to which a proposition is superordinate in a text on memory of those propositions. A study by Duell (cited in Frase, 1975) showed that textual information which is judged to be important is better recalled than unimportant information. A study by McKoon (1977) showed an effect of importance of propositions on the speed and accuracy of sentence verification, and Johnson (1970) found that structural importance of verbal units was positively related to recall. In addition, thematicity is an important concept often included in text linguistic theories, for instance that of Crothers (1979b). These studies suggest that it may be possible to design adjunct questions, taking account of the importance of the information questioned relative to a given topic and reading goal, to manipulate their effects on comprehension and memory.

Experiments especially pertinent to the present one have been reported by Rickards and DiVesta (1974) and by Rickards (1976). They found that thematic or

"conceptual" questions, ones requiring synthesis of particular facts and the relating of them to thematic statements, produced higher performance on a free recall post-test than did "verbatim" questions, ones that referred to single sentences of the text. The present study extended Rickards' work in several important respects: (i) Natural prose was now used, instead of the rather artificially constructed passages of Rickards. (ii) A possible confounding between the verbatim - paraphrase and thematic level variables inherent in that study was now eliminated. (iii) Rickards' dependent variable had six levels, some of which he showed to be substantially correlated. Yet he ran six separate ANOVAs. In the present work this error was avoided, by performing a multivariate analysis of variance (MANOVA). (iv) The conditions in Rickards' study produced only low levels of recall. By contrast, the procedural changes introduced here were designed to bring memory up to what one usually regards as more typical.

Method. The design contained three crossed between-subjects factors: version (Original, Revised); thematicity of adjunct question (Thematic, Non-thematic, Irrelevant adjunct questions); and relation of question form to the form of the corresponding information in the text (Verbatim, Paraphrase). Dependent variates were test frequencies correct and text reading times.

Participants were college students serving as part of a course requirement in introductory psychology. Fifteen subjects were randomly assigned to each of the 2x3x2 = 12 cells of the design. The text passage used was Nobel Prize, the revised version being that of Experiment 3 of project I reported above. (This investigation began well before the "proof" analysis had been conceived.) Thematicity of adjunct questions was determined by the level of the corresponding assertions in the theoretical graph. Each adjunct question following the usual procedure, was inserted in the text directly following the segment that supplied the answer. All adjunct questions called for short answers, this form having previously been shown to yield the largest effects. Post-test questions had the same form, short answer, as adjunct questions. Cued recall was used on the post-test.

All instructions and materials were presented via CRT keyboard display units, in a sequence controlled by the computer. Responses and latencies were likewise recorded. After the initial instructions, sentences of a warm-up passage were presented one at a time on the display screen. Readers proceeded at their own pace,

pressing a button to deliver the next sentence. At appropriate points adjunct questions appeared and participants wrote their answers on a separate sheet. Upon completion of the passage, the post-test was administered, with the order of questions being randomized individually for each participant.

Model and Results. A flow chart model of processing adjunct questions was constructed. The model contained a text encoding component, a question encoding component and a question-answering component. Text version was hypothesized to have its largest effect on the construction processes of the text encoding component, thematicity on the rehearsal processes of text encoding, and the verbatim-paraphrase variable on the answer reconstruction process of the question-answering component.

Partial support was found for the model's distinction between thematic and nonthematic rehearsal processes. Reading times for groups receiving thematic adjunct questions were significantly longer than for groups who received non-thematic adjunct questions. Also, partial support was found for the model's distinction between lexical and semantic reconstructive processes. That is, groups receiving verbatim adjunct questions showed higher average test scores than groups receiving paraphrase adjunct questions.

Text version did not significantly affect criterion test scores. An increase in reading time which could not be accounted for in terms of difference in passage length was noted for the revised text.

Groups receiving irrelevant adjunct questions were hypothesized to have a separate flow of control, omitting text rehearsal. Those groups had both significantly shorter reading times and lower criterion test scores compared to groups receiving thematic adjunct questions. Groups receiving irrelevant adjunct questions performed equally poorly on thematic and non-thematic criterion test questions.

III. Using adjunct outlines in an attempt to assist text memory. (Edward Crothers, assisted by David Williams).

Experiment 7.

This study is probably worth reporting at some length despite the negative results, because the method is rather novel and also because the hypotheses appeared quite plausible at the outset. The hypothesis in

its broadest form was that text comprehension, and hence later memory, can be facilitated by presenting the text in certain unconventional formats that clarify the structure. Here the focus was on formats that supplement the text with an outline, but do so in novel ways. One conceivable drawback of a conventional text-plus-outline format is that presenting the outline separately from the text requires the reader to discover for himself the correspondence between outline lines and text segments. In other words, the internal organization of the outline is exhibited but the external organization, between outline and text, is not. Conversely, a conventional alternative procedure, namely inserting the outline lines one by one as headings directly within the text, poses the opposite problem: The outline-text match is explicated but at the expense of the outline integrity. A novel outline format was sought that would retain the best of both worlds, in the hope of thereby increasing comprehension and memory. At the same time, it was realized that any new format would need also to be quite simple. Otherwise it might introduce familiarization difficulties that could offset the theoretically more efficient layout. A further consideration was the need for control conditions appropriate both to the hypothesis that the new format would be an improvement over the text alone and that it would be more effective than a conventional text-plus-outline format.

Method. To explore alternative possibilities, two different experimental formats, rather than only one, were designed. Also, there were two control formats, one being the text alone and the other the ordinary text-plus-outline format. In order to avoid complications attributable to transfer between formats, each subject saw only one format.

One of the experimental formats, called a run-in outline, is exemplified by the following passage, which is one that was actually used in the experiment. For brevity here only the beginning is reproduced:

"Catastrophe" Theory

Catastrophe theory. — I. Definition. A mathematical theory has recently been developed for describing changes that are abrupt and discontinuous as opposed to smooth and gradual. In everyday language we often use expressions such as "the straw that broke the camel's back" to talk about abrupt changes.

Mathematicians have dubbed this new theory "catastrophe" theory. Here of course the word "catastrophe" does not mean some sort of disaster, but merely any abrupt change.

Catastrophe theory. I. Definition.→ II. Purpose. The purpose for which the theory is intended is to provide both a mathematical theory of, and scientific models of, discontinuous phenomena. Catastrophe theory. I. Definition. II. Purpose.→ A. Purpose in mathematics. Within mathematics the theory offers promise for filling a major gap···

The other experimental format, called an indented outline, is exemplified as follows, using the same initial excerpt:

"Catastrophe" Theory

Catastrophe theory

→I. Definition

A methematical theory has recently been developed for describing changes that are abrupt and discontinuous as opposed to smooth and gradual. In everyday language we often use expressions such as "the straw that broke the camel's back" to talk about abrupt changes. Mathematicians have dubbed this new theory "catastrophe theory." Here of course the word "catastrophe" does not mean some sort of disaster, but merely any abrupt change.

Catastrophe theory

I. Definition

→II. Purpose

The purpose for which the theory is intended is to provide both a mathematical theory of, and scientific models of, discontinuous phenomena.

Catastrophe theory

I. Definition

II. Purpose

A. Purpose in mathematics.

Within mathematics the theory offers promise for filling a major gap...

Note first that these formats have several features in common: Each alternates text with outline, each develops the outline cumulatively, and each uses a pointer to denote the upcoming text topic. Also, of course, the text and outline themselves are identical for both formats, and the outline is not so elaborated as to be confusing regardless of the format. The difference between formats is that the indented format obviously presents a greater typographical separation between text and outline than does the run-in form. Therefore in addition to the original hypothesis relative to a conventional outline, one might also speculate as to which of these two experimental formats will prove most effective. Perhaps it is important to display the outlines in a semi-conventional format that encourages scanning of each outline segment as it occurs. If so, the indented format should be superior to the run-in one. On the other hand, perhaps typographical segregation of the cumulative outline makes it easy for a reader to disregard it, thus undermining the entire purpose of the outline. If so, the run-in format should be superior, because it is less segregated typographically, hence harder to skip. No doubt for either format the repetition inherent in any cumulative outlining will unavoidably tempt a reader to engage in skipping and skimming to some degree, however.

One of the two control conditions was a format control, in that the same text and outline were presented but with the outline in a conventional format; that is, simply typed on a separate page rather than cumulatively alternating with the text. The other control condition controlled for the presence of any outline format, by being simply the text alone.

In order to explore the possibility of an interaction between format effectiveness and text difficulty, a second passage in addition to Catastrophe

Theory was also prepared in the four formats. The other passage was again Nobel Prize. Based on the earlier work as well as on the fact that Catastrophe Theory is clearly more abstract than Nobel Prize, it was again anticipated that Catastrophe Theory would be the more difficult of the two. For pedagogic relevance, it seemed preferable to use passages that were themselves well organized. Hence the revised versions (revisions of the form of Experiment 5 in part I) were used.

The experimental design was entirely between-subjects. Thus there was a group for each of the eight combinations of four format conditions and two passage conditions. Originally it had been planned to run each subject twice with the same format, once on each passage, in order to observe any learning-to-learn effect. Unfortunately, the duration of the passages, especially owing to cumulative outlining, compounded with the fairly lengthy tests made it impractical to run a subject on more than one passage. At least, a minimal amount of practice was provided by the inclusion of a brief demonstration passage, in its appropriate outline format, as part of the initial experimental instructions. In other respects as well, the instructions provided rather complete preparation to the participant. Not only was he instructed carefully about the formats, but also he was warned to expect intensive tests of his memory for the passages.

Passages were presented in dittoed booklets. Participants, again volunteers from the introductory psychology subject pool at the University of Colorado, were run in small groups. Each reader was allowed an experimenter-paced amount of time for each phase of the experiment. The phase times, which were identical for all groups of subjects, had been determined from completion time data collected in a subject-paced pilot study.

Immediately following the reading of a passage, two tests were administered. Because neither passage offered many opportunities for significant inferences, both tests involved primarily paraphrase or verbatim memory. The first test was a detailed fill-in test, querying both thematic and detailed information from the passage. The Catastrophe Theory and Nobel Prize fill-in tests contained 45 and 54 blanks, respectively. The question sequence adhered to the original sentence sequence in the passage. The second test called for a free recall of the outline, thus providing a second opportunity to observe performance differences among

the four format conditions. For a subject in the outline control condition, who had not received an outline in the first place, this second test instead asked him to construct from memory his own outline of the passage.

Results. The data can be summarized briefly. Little of interest was found after ten subjects had been run in each of the eight cells of the design. Hence the study was terminated at that point. More specifically, these data exhibited no trends, let alone statistically significant differences, among the four formats for either of the two passages on either of the two tests (other than the trivial fact that, of course, the text-only subjects fared poorer on the outline test). Mean proportions of errors were astonishingly high on the fill-in test, namely .40 for Catastrophe Theory and .25 for Nobel Prize. On first thought it was suspected that these high error proportions were responsible for the lack of format effects. Three findings argue against this conclusion, however. For one thing, there were no appreciable group differences on the outline recall test, where (as one might expect) the scores were much higher. For another thing, a reanalysis of the fill-in data excluding subjects whose error frequency exceeded an arbitrary cutoff (viz., 29 errors) confirmed the failure to observe format effects.

In retrospect, the negative result, though disappointing, is not without value for the purpose of understanding the possible educational benefits or lack of benefits obtainable from innovative methods of text outlining. The absence of an interaction between format condition and passage, together with the failure to find a facilitative effect for either the high-scoring or the low-scoring subjects, rules out certain otherwise plausible explanations of the negative outcome. The other side of the coin is that several possible explanatives remain. The immediately suggested conclusion, of course, is that the hypotheses are simply incorrect over the whole range of educationally interesting experimental conditions. This may well in fact be correct, though the present results do not warrant such a sweeping verdict. The other possibility is that the subject's motivation, or else his experience with an unfamiliar outline format, or both, was insufficient to produce positive effects. It is true that unpaid subjects in studies of this sort are often observed to be rather unmotivated, and it is also true that a one-hour session hardly amounts to an extensive training session. What is uncertain, however,

is whether or not higher levels of motivation and practice would indeed have produced significant facilitative effects of the novel formats.

IV. Using adjunct connections to assist text comprehension and memory. (Edward Crothers, assisted by David Williams).

Experiment 8.

This more recent project is almost completed. It will be described next here, because it continues the investigation of adjunct aids to text memory. Chronologically, however, this project was a resumption of that line of work, after the deeper theoretical extension summarized in part VII herein had taken place. As the title of this fourth project suggests, now the intended role of adjunct aids was shifted -- from displaying the outline to the more concrete function of displaying the cohesion among the explicit text sentences directly. A significant change in the research means, though not in the ends, is evident here. Neither inferable propositions (as in the revised versions of project I) nor questions (project II) nor abstract outlines (project III) were now inserted. Instead, all inserted adjuncts were inferences of connectives among text sentences. Or to state it more accurately, the only information inserted beyond the original text was certain connectives, though repetitions of corresponding sentences were also inserted in order to clarify just what each connective was connecting. The rationale now for emphasizing these connections is only secondarily the hypothesis that comprehension sometimes suffers from a reader's inability to make the connections. Rather, the primary view is that a reader often is not sufficiently attentive to engage in more than moderately deep processing, to the detriment of later memory.

Another goal was to extend the research to at least a modest number of additional text passages, beyond the two to which the previous experimental work had been confined. A further notable departure from the earlier projects was the close relationship now established between adjunct insertions and post-test questions. The former approach of testing all knowledge of the text was abandoned in favor of testing only for connections. However, this was done with a different twist than in the reading materials themselves. On the test the participant, instead of being given two text propositions and asked to supply their connective, was given one plus the connective and asked to recall the other. It might be added

that testing by recall rather than by yes-no or multiple choice methods poses no serious data scoring problem when, as here, the required responses are no lengthier than one phrase or sentence. And of course recall testing does avoid the complications inherent in response selection methods, especially the effects of guessing strategies and of the particular incorrect alternatives offered.

Method. Four passages were selected and edited for clarity. In addition to the well-known narrative Circle Island, two of the others were descriptive passages of similar lengths, entitled The Aztecs and Pygmies. The fourth was a two paragraph passage, entitled Java, about the island of that name.

By far the major task of this project was the development of a promising format for imparting to the reader the adjunct inferences of the main connections. Clearly it is a long step from deciding to highlight a text's sentential cohesion to devising an efficient, readable format for implementing that decision. The format finally settled on for the experiment was actually the fifth one devised; it may be of interest here to describe the first four as well. The problem was not really one of devising informative formats for displaying coherence; indeed, several readily came to mind. Rather, the problem was the rather commonplace one of finding a format that is not only informative but is also easily understood and used by a reader. Placing a premium on usability entailed that the format idea actually be quite simple and straightforward. But it should also be borne in mind that some "obviously helpful" possibilities turned out to be otherwise when examined more carefully.

The format finally adopted is illustrated below for part of the Java passage.

1. Java is a small island, 2. narrow in shape, and 3. lying south of the equator. 4. Its total land area is 51,000 square miles,

[Therefore 1. Java is a small island.]

or 5. slightly larger than New York state. 6. It has few towns, and 7. most of its people are farmers. Yet 8. it has a population exceeding 45 million, or 9. three times that of New York state. 10. The average population density is more than 800 persons to the square mile,

[Because 4. its total land area is 51,000 square miles and 8. it has

a population exceeding 45 million.]
11. which is impressive. But 12. the density reaches the fantastic figure of 1600 persons to the square mile in the rich farm lands 13. which Java has.
[Therefore, 7. most of its people are farmers.]

Salient features of the format are these: (i) Inferences that enhance the coherence are inserted, and are displayed by being set off on a separate line, using wider margins and enclosing them in square brackets; (ii) Importantly, such inferences are restricted from the class of theoretically allowable ones to sentences in the text itself. Thus what are being inferred are a repetition of a text sentence and a connective from it to the immediately preceding explicit sentence of the text; (iii) This particular format also exhibits the decision to display only the backward-connecting inferences, not the forward-connecting ones. For example, the connection between sentences 4 and 10 is shown after sentence 10, not after sentence 4; (iv) When there is a chain of reasons or causes, the most immediate connection, rather than a remote one, is displayed. (Here immediacy is defined in terms of the causal relations, not in terms of the linear order of text sentences); (v) To avoid belaboring the obvious, connections between adjacent sentences are not shown; (vi) All logical and semantic connections among non-adjacent sentences are shown, including both necessary and probable inferences of both antecedents and consequents; (vii) Depending on the type of connection inferred, the inferred sentence is prefaced by one of the four connective phrases "because", "partly because", "therefore", or "partly therefore"; (viii) The stated text itself is in no way altered to make the connection more explicit, because to do so would detract from the readability. In particular, syntactically reduced surface constructions (clauses, or even phrases) are not raised to sentences despite the fact that all inferred constructions are themselves sentences. An example is clause 13, which is left as which Java has although the underlying Java has rich farmlands is of course what connects to the inferred Therefore 7. most of its people are farmers; (ix) For texts in general, the connected units of course need not be restricted to single sentences. Either the explicit half of the connection, the inferred half, or both, can itself be a multi-sentence subtext. The only example shown in the excerpt is the conjunction of sentences 4 and 8 inferred as a reason for sentence 10.

In passing, it is also interesting to note that connections of the kinds exemplified here are more frequently inferable in texts than had been anticipated. The excerpt shown here is too brief to suggest this observation, but in fact this Java passage of 23 sentences allowed 7 inferred sentences, and the ratio was even higher in the three other passages analyzed. Therefore, despite the fact that except for Circle Island no particular effort was made to select passages rich in inferences, it was emphatically not the case that Circle Island was the only one of the four passages to invite many inferences for cohesion. This outcome suggests another research direction, which however must be left for the future. Conceivably, many texts can be reorganized by various routine techniques (permutation, addition, deletion, etc.) to reduce the number of these nonadjacent connections. In order to do so in a non-piecemeal fashion, the complete interproposition graph would have to be inferred.

The first four attempts to design a promising format will now be summarized briefly:

(i) The first idea also included in-text sentence indexing and explicit statement of connections, but without explicitly displaying the inferred repetition of another text sentence. For example, in one variant of this the first sentence of Java was written as

1. Java is a small island ( $\rightarrow$ BECAUSE 4).

Five variants of this were prepared for each of the four passages, namely the variants including forward, backward, antecedent, consequent, or all inferences. The latter was too inferentially dense to be read smoothly. Among the other forms, both theoretical and speculative considerations suggested that the backward variant would be the one most likely to assist comprehension.

(ii) The second attempt was based on a very different principle than the others, and intended for a different purpose. It involved what might be called "conceptual" or "topical" indexing. No proposition, or even connective, inferences were displayed at all. Instead, each main "thread" or path of the text was indexed with a different integer, assigning all undeveloped, parenthetical statements the integer "0". The intention was to thereby assist a reader in identifying topic shifts, resumptions, convergences, and so

on. The rationale for this clause-numbering scheme was that many texts interweave their topics in a complex fashion that is not immediately recognized by a reader. Again using the first paragraph of Java for illustrative purposes, it was rewritten as:

1. Java is a small island, 0. narrow in shape 0. and lying south of the equator. 1. Its total land area is 51,000 square miles, 2. or slightly larger than New York state. 1. It has a few towns, 3. and most of its people are farmers. 1. Yet it has a population exceeding 45 million, 2. or three times that of New York state. 1. The average population density is more than 800 persons to the square mile, 0. which is impressive. 1. But the density reaches the fantastic figure of 1600 persons to the square mile in the rich farm lands 3. which Java has.

The main reason for abandoning this format idea was that it is actually more appropriate to a different purpose from coherence explication, namely subtopic identification. Indeed, if that objective were to be pursued then some format of this sort might be quite useful. Further improvement would be needed, however, because the numbering scheme proved unworkably complex (i.e., multiple integers assigned to the same sentence) when applied to a more intricately interwoven text such as Circle Island.

(iii) The third and fourth ideas resumed the attempt to devise a format for displaying inferable coherence. The third was a simple adaptation of the standard typographical notation for footnotes:

1. Java is a small island, 2. narrow in shape, 3. lying south of the equator. 4. Its total land area is 51,000 square miles[1], or 5. slightly larger than New York state[4]. 6. It has few towns, and 7. most of its people are farmers.[6] Yet 8. it has a population exceeding 45 million, or 9. three times that of New York state.[8] 10. The average population density is more than 800 persons to the square mile, [4,8] 11. which is impressive. 12. But the density reaches the fantastic figure of 1600 persons to the square mile in the rich farm lands 13. which Java has.[7]

Again, the variant shown is the one that exhibits the backward connections. It would seem that this time too much perspicuity has been sacrificed for the sake of conciseness. In order to recover the referenced connec-

tions a reader must engage in much backtracking, which is hardly assisted by the dual role of in-text integers -- as both sentence numbers and footnote numbers.

To avoid this latter objection, as the fourth attempt the referring superscripts were simply moved to the margins:

> 1. Java is a small island, 2. narrow in shape, and 3. lying south of the equator. 4. Its total land area is 51,000 square miles, or 5. slightly larger than New York state. 6. It has few towns, and 7. most of its people are farmers. Yet 8. it has a population exceeding 45 million, or 9. three times that of New York state. 10. The average population density is more than 800 persons to the square mile, 11. which is impressive. But 12. the density reaches the fantastic figure of 1600 persons to the square mile in the rich farm lands 13. which Java has.

Of all the versions, this one least disrupts the original text. No information about any backward connection -- its number, connective, or wording -- is inserted within the text itself. However, this idea was also judged too compact and nonredundant to encourage easy recovery of the backward connections. Hence the fifth version was developed and adopted, as already discussed.

It remained also to prepare a set of comprehension test questions from each of the four passages. For the reason discussed at the outset of this project, the test questions required response production rather than selection. And of course it was emphasized that the questions were to be answered on the basis of the information in the passages, rather than from any other knowledge the reader might have of the subject matter. The decision was made to test comprehension by a reader's ability to supply queried immediate antecedents (reasons, causes) for stated consequents (conclusions, effects). Therefore, since the backward-forward classification of inferences is orthogonal to the antecedent-consequent classification, some test questions required answers from the inserted, bracketed repetitions. Further, because one would want to test understanding of major antecedent-consequent connections and because the set of such connections contains the set of ones highlighted by the format (as well as containing the

connections among adjacent sentences), the set of queried connections included as a proper subset the ones explicated by the format. Depending on the type of connection, the question form was any one of the following:

1. What is one reason why (consequent)?
2. What is one possible reason why (consequent)?
3. What is one reason from which we can conclude that (consequent)?
4. What is one possible reason from which we can conclude that (consequent)?
5. What is one reason which enabled (consequent)?
6. What is one specific statement offered to support the generalization that (consequent)?
7. What is one generalization offered to support the specific statement that (consequent)?

Sometimes questions of the form What is a second reason why (consequent)? etc. were also included. Chiefly, this was done to ensure that nonadjacent antecedents were tested in those cases where the passage gave both an adjacent and a nonadjacent antecedent. The final versions of the question sets, after several questions had been deleted in order to reduce the duration of the experiment, comprised 7 questions for Java, 11 for Pygmies, 12 for Circle Island, and 15 for Aztecs, making 45 questions in all. Finally, each of the four sets of questions was carefully sequenced so that no question gave away an answer to a later question.

Turning next to the procedure, the central comparison was against a control group who received the same texts in the conventional format. Instructions for the experimental and control groups were nearly identical to one another except that the control instructions made no mention of format. The experimental group was told that the experiment was about how reading comprehension depends on the format of the text. Readers were given a sample text, Java, in the experimental format and instructed on the major and some minor features of the format. The control group was told that the experiment was about what is remembered after reading a short article. They were given the same sample text as was given to the experimental group but in an ordinary format. Both groups were informed that there would be a written test of memory and comprehension after each of the three articles was read. Both groups were given two example questions, without answers,

that were relevant to the sample text. Also, both groups were told to answer the questions in the order they appeared in the text and not to skip around or retrace to change answers to previous questions. This instruction was given since some of the later questions might give away answers to some of the preceding questions. Both groups were instructed to read each article twice before turning in the article and taking the written test.

Experimental sessions were conducted with groups of from three to six college student participants. Subjects were randomly assigned to the experimental or control group, with twenty subjects in each group. The order in which a subject received the three texts was randomly determined. In turn, this determined the order of the three tests, because a test for a particular passage was given immediately after the participant turned in that text.

The entire session was self-paced. As soon as the participant finished one task, he or she turned it in to the experimenter and then received the next task. The time each participant took to read each article, to take each test, and to read the instructions was recorded. Although this time measurement was crude, it did allow a rough analysis of the time differences among groups, passages, and texts. Times required to complete the experiment ranged from about 40 to 60 minutes.

Results. Data from Circle Island and Pygmies have undergone preliminary scoring, with Aztecs scheduled to be scored next. The preliminary scoring is a classification of each response merely as correct or as incorrect, disregarding the exact type of error. An eight-category taxonomy appropriate to the errors observed has been formulated, but it has not yet been applied to the data. As always, the motivation for a refined data tabulation is not that the correct-incorrect tally gives an inaccurate count of the frequencies correct, but rather that it gives only a partial picture of what was remembered. For the present data only 3% of the errors were omissions; the remainder in this recall task were hardly random guesses.

The preliminary frequency tabulations for Circle Island and Pygmies were converted to percentages of correct responses, which were in turn subjected to an

ANOVA. The percentages and ANOVA results appear in Tables 4 and 5, respectively. All three effects -- group (experimental or control), text (Circle Island or Pygmies), and group x text were statistically significant. As Table 4 shows, the direction of the interaction is a much greater facilitation by adjunct connective insertions for Pygmies than for Circle Island.

The reading time data, which have been averaged but not yet tested for statistical significance, appear in Table 6. Group means are given separately for each of the two passages and, within each passage, for the reading phase completion time and the test completion time. Not unexpectedly, the experimental group, having the insertions to read in addition to the normal text, took longer during the reading phase than did the control group. The average times summed over the two passages were 10.05 and 7.20 minutes, respectively. However, the group comparison reversed itself on the test completion times, which averaged 11.25 and 16.83 minutes, respectively, over both passages. Including both text and test, then, the total time averaged a little longer for the control group (24.05 min.) than for the experimental group (21.30 min.). Further, there is a hint of an interaction between passage and group; using total time (text plus test) as the measure, the experimental group averaged 0.65 minutes faster than the control group on Circle Island but 2.10 minutes faster on Pygmies.

Tentative conclusions based on the data analyses conducted to this point are as follows. Although the experimental group predictably took longer than the controls to read the passages, this was offset or even more than offset by the former's responding faster than the latter on the test. For educational purposes, the dependent variable of greatest practical interest would seem to be the total time, on which measure the experimentals were at least as rapid as the controls. In sum, then, the insertion of adjunct connectives produced significantly higher percentages correct on the post-test, while not increasing the reader's average time for both reading and test together. And the direction of interaction between group and passage was the same for the total time data as for the choice data. That is, by both measures the superiority of experimental group over control group was greater for Pygmies than for Circle Island.

Table 4

Mean Percentages Correct in Adjunct Connectives Experiment

| | | Passage | | |
|---|---|---|---|---|
| | | Circle Island (12 items) | Pygmies (11 items) | Total |
| Group | Experimental | 58.3 | 52.3 | 55.3 |
| | Control | 50.0 | 26.8 | 38.4 |
| | Total | 54.2 | 39.5 | 46.8 |

Table 5

Analysis of Variance Based on Table 4

ANOVA

| Source | SS | df | MS | F | P |
|---|---|---|---|---|---|
| Groups | 5693.6 | 1 | 5693.6 | 17.2 | $p<.001$ |
| Subjects | 12546.1 | 38 | 330.2 | | |
| Passages | 4276.4 | 1 | 4276.4 | 20.3 | $p<.001$ |
| Grp. x Pass. | 1462.9 | 1 | 1462.9 | 6.9 | $.01<p<.025$ |
| Residual | 8006.7 | 38 | 210.7 | | |
| Total | 31985.7 | 79 | | | |

Table 6

Mean Latencies (Min.) in Adjunct Connectives Experiment

| | | Passage | | | | | |
|---|---|---|---|---|---|---|---|
| | | Circle Island | | | Pygmies | | |
| | | Text | Test | Total | Text | Test | Total |
| Group | Experimental | 5.15 | 5.75 | 10.90 | 4.90 | 5.50 | 10.40 |
| | Control | 3.40 | 8.15 | 11.55 | 3.80 | 8.70 | 12.50 |

V. Text processing issues examined using the Circle Island narrative (Edward Crothers, assisted by David Williams, Experiments 9-12; David Williams, Experiments 13-14).

Now the deeper, more inferentially oriented theoretical description based on parallelism (project VII) was applied to investigate inferencing during comprehension. Processing hypotheses were now at issue, instead of any further efforts to develop more effective methods of reading instruction. One underlying theme of the work was methodological: to devise brief comprehension tasks that can be inserted into the reading material and then to compare performance against memory data on a post-test. The other theme was structural: to formulate and test hypotheses about the relative difficulty of comprehending and remembering different kinds of text connections identified by the theory. Three pilot studies and then one major experiment have been completed, with a second in progress at the time of this writing.

Experiment 9.

An exploratory pilot study, intended to begin developing the materials and the techniques of data analysis, initiated the project. The aim was to investigate story understanding using two tasks, a fairly standard task of arranging text sentences into a coherent story followed by a specially constructed task of selecting which text sentence was an antecedent (cause, reason) for a certain other text sentence. It was expected that participants' orderings which conformed most closely to the theoretically derived coherent sequencing would be associated with highest post-test scores. Secondly, certain kinds of post-test connections were expected to be easier to recognize than others.

Method. Twelve subjects, again from introductory psychology, participated in this pilot study. Individual sentences of the story, or in some instances groups of two or three adjacent sentences, were typed on 3"x5" cards. Then the deck of cards was scrambled and handed to the participant. His task was to put the cards in a cohesive, natural paragraph sequence, proceeding at his own pace.

Upon completing this task, which took only a few minutes, he was given a test designed to indicate which inferences he had drawn. The test contained twelve items, each constructed as follows. First came a Circle Island sentence designated as a conclusion. Then came

three additional sentences of the text, each of which either was or was not an antecedent for that conclusion. For each of these latter three, the participant was to indicate whether or not he understood it to be an antecedent for the stated consequent. Items were constructed so that some of the candidates were in fact antecedents, and were further divided into antecedents that had appeared near their consequent (in the original passage itself, not in the produced ordering) on the one hand, and remote antecedents on the other hand. The other candidates were not in fact antecedents of the particular consequent given. Instead, some were not causally related to it at all; others were related but in the reverse direction, as consequents themselves rather than as antecedents. An example of a test item is:

> Effect or result: The senate allowed only a small canal to be built.
>
> Which of the following was a cause or reason for the above result?
>
> a. The canal was so shallow that no water would flow into it.
>
> WAS WAS NOT
>
> b. The senate consented to build a canal across the island.
>
> WAS WAS NOT
>
> c. All senators are ranchers.
>
> WAS WAS NOT

In this example the correct answer is WAS NOT to alternatives (a) and (b) and WAS to alternative (c).

Results. The data are of interest in three respects: performance on each of the two tests separately as well as the correlation between the two tests. Not all of the pertinent analyses have been completed, but probably the most revealing ones have been done. In actual practice, the data analysis were carried out only to the extent deemed necessary to suggest the next study.

Analysis of the sequencing data required some planning, because it was not immediately obvious just what would be an appropriate dependent variable. Our interest lay in the "distances" among sentences in a particular participant's protocol; hence some measure of clustering or proximity seemed suitable. However, the reader arranges the deck only once, and hence methods used to study the

growth of memory structure over multiple recall trials are ruled out here. Also, our interest was more in comparing the produced sequence against some standard, some theoretically highly coherent sequence, than in evaluating the produced ordering by itself. (The theoretically selected standard is more properly called an "ideal" sequence rather than a "correct" sequence. To call something "correct" presupposes a criterion. By the criterion of congruency with a theoretical ordering, a protocol that perfectly reproduces it is correct. Nevertheless, by other criteria such as performance on the post-test, it is an empirical question whether or not such a so-called "correct" ordering is in fact the one that is associated with the highest mean performance.) Two, rather than only one, ideal orderings were derived for comparison against the obtained orderings. The theoretical orderings might be called a "presupposition ordering" and a "premise ordering", respectively. The difference is that in a presupposition ordering (which, it is worth noting, approximately corresponds to the original text itself) the sequence of statements reflects the sequence of presupposed subtopics of Circle Island. Thus, all sentences on the island's governmental principles occur in the same paragraph regardless of how their consequents fan out into scattered parts of the sequel. By contrast, in a premise ordering the distance, measured in terms of number of intervening sentences, between antecedents and their respective consequents is minimized. Taking the same example, the background paragraph on the island's governmental principles and apparatus is split into two halves. The one half serves as background for narrating what the government did about the canal issue, in adherence to its customary principles. Hence that half is placed just before that part of the narrative. But the other half contains the seeds of the government's defiance, in the case of the canal issue, of its other customary principles. Likewise, this second part of the background is placed directly before the particular consequents it led to in the canal issue instance.

Separately for each of the two ideals, the measure of comparison to the produced ordering was calculated as follows. The objective was to compare the ideal matrix of inter-sentence distances against the observed matrix. This was done by getting a difference between distances; for each pair of sentences the theoretical distance was subtracted from the ideal. The difference was then squared and the squares were summed over all pairs, giving $\Sigma d^2$ as the measure of comparison between the two orderings for a given participant. Yet to be done is a rank correlation coefficient instead of a distance measure. Of course, the $\Sigma d^2$ measure does not

tell us much about the comparison of ranks, at least not in any immediately obvious way. For example, a pair placed in positions 1 and 2 by a subject as against positions 9 and 10 in the ideal yields a value for $d^2$ of $[(2-1)-(10-9)]^2=0$.

The outcome was evenly divided. Six of the twelve subjects yielded lower $\Sigma d^2$ when calculated from the presupposition order and the remaining six yielded lower sums for the premise order. This outcome is quite inconclusive, not only because of the even split, but also because of the possibility that perhaps some third order, one reconstructed post hoc statistically from the data rather than a priori from the theory, would more closely approximate the obtained data.

A better understanding of this point has not yet been achieved. A first step would be to get some idea of what is a "large" value of $\Sigma d^2$. We might begin by averaging this quantity over pairs, that is, by dividing it by 136 since there were $\binom{17}{2}=136$ pairs of sentences. Then the square root of this gives the mean d value. If the sequencing is done randomly this should equal 8.5, which is itself the mean between its possible extremes of 1 and 16. Using the presupposition ordering for illustrative purposes, eleven of the twelve individuals yielded values below 8.5. However, for only seven of the twelve was the value below 6.0, so the correspondence was only moderately high. The reason for this outcome is still being sought. One possibility, which is amenable to testing, lies in the text structure itself. Roughly speaking, it could be said that the ideal orderings are "tight" within paragraphs but "loose" between paragraphs. Suppose a protocol closely approximated the ideal within-paragraph ordering but deviated from the between-paragraph sequence. It could still produce a fairly high distance difference, because most sentence pairs are not in the same paragraph.

Despite the fair degree of departure from either ideal ordering, scores on the inference post-test were generally high. The overall proportion of "hits" (correct recognitions of antecedents) was 0.82, against only 0.15 for "false alarms" (spurious identifications of nonantecedents as being antecedents). It is not known how much of this can be attributed to inferencing in the course of arranging the sequence, though the next analysis indirectly suggest that some can be. However, it is also quite conceivable that high scores in the post-test may have resulted from drawing inferences while taking the test itself. This issue was a continuing concern in this series of studies, and will be raised again repeatedly.

An interesting aspect of this first pilot study is the possibility of a correlation between performance on the sequencing and inferencing tasks. If readers do indeed engage in inference-making in order to relate sentences to one another and thereby decide their ordering, and further if many of these are the same inferences as those called for later on the inference test, then inferential skill on the one test should be correlated with skill on the other test. Of course, the correlation would be attenuated if participants postponed their inferencing until it was directly required, namely on the post-test. Unfortunately for the purpose of this hypothesis, the sequencing task does not measure inferencing directly. However, a reasonable assumption is that the closer the person's ordering approaches the one or the other ideal ordering (especially the ideal premise ordering), the more has he drawn the pertinent inferences. Since the most noteworthy inferences are indeed embodied in the inference test items, it follows that the $\Sigma d^2$ discrepancy measure should correlate negatively with proficiency on the inference test. Using the number of hits minus the number of false alarms as the score on the inference test, the obtained correlations with $\Sigma d^2$ were -0.35 and -0.63 for the presupposition and premise orders, respectively. The latter is significant ($p < .05$) but the former is not; $t=1.12$ and $-2.57$, with 10 d.f., respectively.

Despite this confirmation of an hypothesis, the method of this study is inherently limited in what is could ever reveal about ongoing inferencing during reading comprehension. As mentioned, there is ambiguity as to when, on the first task or on the post-test, a reader engages in the inferences of interest. A further source of dissatisfaction with the method was that the sentence sequencing task taps a reader's inferencing skills only indirectly. The next three studies abandoned the initial sequencing task in favor of other tests of ongoing comprehension, as well as adding several refinements to the design. The last studies resume use of the sequencing task.

Experiment 10.

The first improvement on the initial pilot study was to devise, after some trial and error, a more direct way of assessing on-line inferencing during reading. Of course, a range of methods have been reported in the literature, including ones involving monitoring tasks and latency data. Latency is widely heralded as a sensitive measure. But the usual monitoring tasks appear to induce an abnormal reading situation. Therefore a fairly natural alternative procedure was devised.

Method. Again, single sentences or groups of a few successive sentences were displayed one at a time. But now (i) they were in correct order, and (ii) five of the sentences had been transformed into alternatives that were either implausible or, in some cases, contradictory in the context of the preceding sentences. As an example of a contradictory continuation, the two antecedent sentences The ranchers are more prosperous than the farmers and The senators consist of the richest islanders were followed eventually by the "consequent" All of the senators are farmers. An alert participant should, of course, reject this continuation, recognizing that it contradicts the more probable consequent All of the senators are ranchers. Because the study sought to concentrate on the most interesting inferences, only five items from the twelve in the first study were retained. Each of the five was transformed as just exemplified; hence for each the correct answer was "no" (not a plausible continuation). Apart from these five each other sentence of the original text was preserved, and hence upon encountering each the correct answer was "yes". Of course, the interesting data were expected to come from the five implausible, "no" items, the "yes" items being easily recognizable as such.

The study was programmed in the DTES (Display Terminal Equipment System) package of CLIPR programs. Upon seeing a sentence, the participant responded either "yes" (plausible) or "no" (implausible), indicating his response by an appropriate button press. Correct acceptances were followed by the next sentence. Incorrect rejections (false alarms) were followed by the next sentence together with a statement to the effect that the preceding sentence was plausible. Incorrect acceptances were followed by a substitution of what the correct sentence should have been (e.g., All of the senators are ranchers). The feedback was designed to minimize the perpetuation of errors over the course of the passage. Choice latencies were recorded. After one pass through the passage, there followed the same twelve-item post-test of inferences as described for the preceding study.

Results. After running only a few persons, it became clear that the procedure needed to be modified. Participants did find it inordinately difficult to spot the five implausible sentences. But, judging from their post-experimental comments, some of the difficulty arose from an uninteresting source. An unfortunate fact about the passage was that most of the five "no" items were inherently only implausible, but not outright contradictory, in context. Hence the

participant faced an unintended dilemma of deciding how implausible an offered continuation had to be before he rejected it. The study was immediately modified, as described next.

Experiment 11.

Method. Now the "yes"-"no" task was replaced by a forced-choice task. The less inferentially interesting continuations, those whose correct answers had been "yes" in the preceding study, were now merely presented one at a time in correct order without requiring any overt response. At each of the five interesting points, both the original, consistent continuation and the transformed continuation offered in the preceding study were presented concurrently. The participant's task was to choose which of the two he considered the more plausible one in context. No post-response feedback was given.

Directly thereafter, there occurred the inference post-test, refined from the preceding studies. Now the alternatives were recomposed so that every one was in fact either a consequent or antecedent of the stated conclusion (rather than some being merely postcedents or precedents, as before). Again, the only group of subjects run was a group whose task was to decide whether or not each of the four alternatives offered with the conclusion was or was not an antecedent for that particular conclusion. Again, all four alternatives were explicit sentences of the text. Besides the direction of the connection between conclusion and alternative, the other variable was the "directness" of the connection. Directness was defined in terms of the relative number of propositions intervening in the surface text between antecedent and consequent, rather than in terms of the semantic or logical immediacy of the inference. Specifically, any given item had four alternatives: a direct antecedent, a remote antecedent, a direct consequent, and a remote consequent. Another feature was that five of the inference test items corresponded to the five critical sentences of the forced-choice task, enabling a comparison between performance on the two tasks. The procedure was programmed and run on DTES (Display Terminal Experimental System), thereby enabling response latency measurement.

Results. The forced-choice task turned out to be quite easy. Four of the five items were rarely missed; however, the other one was missed by 12 of the 19 subjects. The latter item was unique in that the incorrect alternative was a pragmatic contrary of the original text. The original sentence was All the senators are ranchers and the pragmatic contrary was

Most of the senators were ranchers. Readers who chose All took longer to decide than did those who chose Most. The respective mean response latencies were 24.6 and 12.9 sec., $t = 3.12, df = 17, p<.01$. This suggests that more inference-making was required to choose the correct alternative. Analysis of the corresponding item of the inference test revealed no differences, either in errors or in latencies, between participants who chose All and those who chose Most.

Most of the data analyses pertain to the inference post-test. Taking error frequency as dependent variable, neither the effect of directness nor the interaction of directness with relation was significant. However, among the choice alternatives significantly more errors were made on antecedents than on consequents ($F = 21.5$, $df = 1, p<.001$). In other words, participants more often correctly identified a consequent alternative as not being an antecedent of the test sentence in question (WAS NOT response) than they correctly identified an antecedent as such (WAS response). However, it is obviously premature to interpret this outcome as implying that consequents are easier to recognize than antecedents, relative to a test statement. Alternatively, the outcome might merely reflect the asymmetry in the design: readers were asked only to identify antecedents.

Analysis of the response latencies turned out not to be straightforward. Often in response latency studies, the response latency is averaged within each subject's data for each of the conditions (in our case, direct antecedents, remote antecedents, direct consequents, and remote consequents). These averages are then used as the dependent measure in an ANOVA. However, if the number of observations going into each mean depends on the condition being averaged, this general procedure is inappropriate and may even be misleading. Unfortunately, this is the case here. Because each mean would be found for correct responses or incorrect responses, and moreover the relation (antecedent vs. consequent) did have a main effect on error rates, hence the number of observations for each mean is hardly random. Rather, it is directly related to relation. Hence a thorough analysis of the response latencies requires using more cumbersome multiple regression techniques. Such an analysis allows the "independent" variables to be non-orthogonal, so that averaging of data is not necessary. To date only the response latencies associated with correct responses have been thus analyzed.

For the directness variable, it was of course expected that a remote antecedent would take longer to identify correctly as being a cause or a reason than would a direct antecedent. In fact, no main effect of directness or of relation was found. However, the interaction of the two variables approached significance ($F=2.3$ $df=1,648$, $p<.25$). The mean response latencies of remote antecedents, direct antecedents, direct consequents, and remote consequents were, respectively, 13.5, 11.7, 13.2, and 12.0 sec. Even though, by the obtained F, this interaction cannot be said to be a rare event, it can be deemed rare enough to encourage continuing this line of investigation. For one thing, the analysis was based on only 19 subjects. For another, both the between-subjects variance and the within-subjects variance were quite large. Hence the statistical test does not have much power. It should also be pointed out that the within-subjects variance exceeded the between-subjects variance. Probably the reason is that the variance between items was not taken into account in the analyses to date, as it will be in the future.

This interaction is of some interest, because it is an X-interaction (i.e., an interaction of two variables neither of whose main effect was significant). What this would imply, if it were replicated in a larger scale study, is that readers do not simply search their memories to ascertain the relation between two sentences. To expand on this point, recall what this X-interaction means in the context of the present pilot study. A direct antecedent was more quickly identified as an antecedent that was a remote antecedent. This is a consistent with a simple search hypothesis. On the other hand, a remote consequent was more quickly identified as not being an antecedent thanwas a direct consequent. This implies that there is something other than a simple search that "gives away" the remote consequent. In studying what this something is, or more precisely, in studying how subjects manage such a complex strategy, it would be profitable to use control groups that assist in identifying the source of this phenomenon.

Experiment 12.

The purpose of this larger experiment was to replicate and extend the previous study. Additional experimental and control groups were added in order to compare the ease of inferring antecedents and consequents and to pursue the suggested X-interaction between directness and relation of the post-test choice alternative to the post-test statement.

Method. During the initial phase, text reading, each participant's task was the same as before, namely to make a forced-choice at five points in the sentence sequence as to which of two sentence alternatives would be the more plausible continuation. The testing phase, however, was modified. Test items themselves basically had the same properties as in the preceding experiment. The items were edited, however, and reduced from twelve to six in order to obtain a set more homogeneous in the detailed inference type and in sentence length. The display format for this revised inference test resembled that of a sentence verification task:

WAS
(test sentence)
A CAUSE OR REASON FOR
(key sentence)?

Press button A if "Yes"
Press button E if "No"

Again, each test sentence and each key sentence was taken directly from the story. Each of the six items consisted of four successive displays of the form shown here, holding constant the key sentence and varying the test sentence. The latter were selected to meet the constraint that each set of four contained one relatively direct antecedent with respect to the key sentence, and likewise one relatively remote antecedent, one direct consequent, and one remote consequent. (Recall that in these studies a textual, not a semantic, definition of directness is being adopted, namely number of intervening sentences in the original text.)

Three experimental groups were run, each consisting of 21 volunteers serving in partial fulfillment of a course requirement in introductory psychology. Assignment of participants to groups was random. The first group was a replication, except for the item revisions mentioned above, of the group in the previous study. This group (Antec) was on the post-test always probed about antecedent relations. Conversely, the second group (Conseq) was always probed for consequents. Schematically, this group had the following format:

WAS
(test sentence)
AN EFFECT OR RESULT OF
(key sentence)?

Press Button A if "Yes"
Press Button E if "No"

Unlike either of these two groups, the third group of subjects (Ant /Cons) were probed for antecedents on some items, consequents on others. The decision as to which to probe was made randomly, subject to the constraint that for each key sentence on exactly two of the four test sentences the affirmative response was correct. To illustrate how this was done, the following table lists the $\binom{4}{2}=6$ admissible combinations of text sentence type and probe question type. One of these six was randomly chosen for each item; that is, each key sentence and its four test sentences. The test sentence types are DA (direct antecedent), RA (remote antecedent), etc. The probe question types are A (antecedent) and C (consequent), shown as entries in the table. An asterisk identifies each combination for which a "yes" response is correct.

| Test Sentence Type | | | |
|---|---|---|---|
| DA | RA | DC | RC |
| A* | A* | A | A |
| C | C | C* | C* |
| A* | C | C* | A |
| C | A* | A | C* |
| C | A* | C* | A |
| A* | C | A | C* |

Note that the randomizing does not necessarily balance over participants or over items. For example, there is no constraint that for a given person the response "yes" be correct to exactly three of the six DA test sentences.

Of course, the six key sentences and associated twenty-four test sentences were identical for the three groups of participants. For each group the presentation sequence was randomized in two steps. First the six key sentences were randomly ordered. Then the four test sentences with a particular key sentence were presented consecutively, also in random order. The A button (leftmost button) was always used for affirmative responses, the E button (rightmost button) for negative responses.

The initial instructions for the inference posttest were neutral with respect to antecedent or consequent relations, except that the demonstration examples were appropriate to the subject's task (e.g., consequents probed in the examples shown group Conseq). Two examples were presented, one having the affirmative response as correct and the other having the negative.

The negative example was accompanied by a caution against answering a question "backwards", (i.e., confusing key sentence with test sentence). Post-test questions were presented in immediate succession with no appreciable inter-item interval. The entire experiment was controlled by the Display Terminal Equipment System (DTES) program executed on the Xerox Sigma III computer in the CLIPR lab.

Results. Data tabulations and some significance tests using latency and choice data were calculated according to the main variables in the design, without pursuing further, more detailed, analyses such as individual item breakdowns.

Mean reading times on the initial phase (reading, interspersed with five forced-choice items) were not significantly different among the three groups; $F(2,66)=1.84$, $p>.10$. However, the mean over participants tended to be slightly higher for the Antecedent group, namely 252.9 sec. compared to 221.1 sec. and 212.3 sec. for the Consequent and Antecedent/Consequent groups, respectively. The only correlation coefficient computed between this reading time measure and the frequency correct on the post-test was a single Pearson r calculated from all three groups' data combined. No correlation was found ($r=.007$), in agreement with the usual finding though separate correlations for the three groups still need to be calculated.

On the five-item forced-choice test the mean frequencies correct were 4.14, 3.76, and 3.81 for the Antecedent, Consequent, and Antecedent/Consequent groups, respectively. Correlation coefficients between scores on this inference test and scores on the inference post-test have not yet been calculated.

Table 7 presents a comparison of the three groups on the six-item inference post-test. From the table it can be seen that group Antecedent surpassed the other two groups on all four test item types. However, for some reason the Antecedent/Consequent group was not intermediate between the other two groups. At any rate, the suggestion is that readers are better able to decide whether or not a test sentence was in the text an antecedent of a key sentence that they are able to decide whether or not it was a consequent. Initially it had been hypothesized that readers tend more to associate a sentence with its consequents than with its antecedents. Hence, the direction of the actual finding was quite unexpected.

TABLE 7

Percent Correct On Inference Post-Test

| Group | Test Sentence Type | | | | |
|---|---|---|---|---|---|
| | DA | RA | DC | RC | Mean |
| Antec. | 81.0 | 65.9 | 70.6 | 74.6 | 73.0 |
| Conseq. | 50.8 | 63.5 | 62.7 | 57.1 | 58.6 |
| Ant./Cons. | 54.0 | 53.2 | 62.7 | 60.3 | 57.6 |
| Mean | 61.9 | 60.9 | 65.3 | 64.0 | 63.0 |

The other main conclusion from the table is that there was no overall effect of test item type. Or to put it more accurately, there was no effect of directness (DA, DC compared to RA, RC), though the possibility of a slight effect of test sentence direction (DA, RA compared to DC, RC) cannot yet be ruled out. The apparent absence of an effect of directness was indeed surprising. Additional data analyses are planned in seeking an explanation of this outcome. Separate tallies for individual items, among which there is considerable variation in response patterns, might help explain the overall absence of an effect of directness.

Latencies of correct responses, shown in Table 8, appear to bear out the trends found in Table 7 for tne percentages correct.

One alternative way of classifying the test sentence types aids in understanding the findings already reported. Instead of classifying the test sentence types as was done in Tables 7 and 8, we can partition them on the basis of whether the correct response is "yes" or whether it is "no". The "yes" cases, of course, are just the items for which the test sentence was an antecedent (consequent) when antecedence (consequence) was queried, namely the DA and RA items for the Antecedent group and the DC and RC items for the Consequent group. For the third group the reclassification depends on the randomization. The outcome for the choice data is shown in Table 9, with corresponding latencies shown in Table 10.

Striking in Table 9 is the suggestion of an interaction between directness and correct response. Direct test sentences are more often responded to correctly when they do in fact exhibit the relation queried than when they do not (72.4% vs. 54.8%). Oppositely, remote test sentences are easier when they do not bear the queried relationship than when they do (68.2% vs. 56.8%). This suggests an interesting conclusion: participants tended to be more influenced by the directness of the relationship than by the direction. They were somewhat inclined to respond "yes" to D test sentences and "no" to R ones, mistaking directness per se for direct antecedence or direct consequence. Consistent with this, Table 10 adds the observation that even when only correct responses are considered participants were more rapid on D-yes and R-no relations than on D-no and R-yes ones.

With respect to latencies of correct responses, was the X-interaction found in the preceding experiment replicated? Yes, it was; the interaction between response ("yes" vs. "no") and directness (direct vs. remote)

Table 8

Mean Latency (Sec.) of Correct Responses on Inference Post-Test

| Group | Test Sentence Type | | | | |
|---|---|---|---|---|---|
| | DA | RA | DC | RC | Mean |
| Antec. | 12.3 | 13.3 | 12.2 | 11.0 | 12.2 |
| Conseq. | 16.8 | 12.8 | 11.9 | 14.7 | 14.1 |
| Ant./Cons. | 13.1 | 13.2 | 11.8 | 12.7 | 12.7 |
| Mean | 14.1 | 13.1 | 12.0 | 12.8 | 13.0 |

Table 9

Percent Correct on Inference Post-Test,
Classified By Correct Response

| Group | Test Sentence Type | | | | |
|---|---|---|---|---|---|
| | D-Yes | D-No | R-Yes | R-No | Mean |
| Antec. | 81.0 | 70.6 | 65.9 | 74.6 | 73.0 |
| Conseq. | 62.7 | 50.8 | 57.1 | 63.5 | 58.6 |
| Ant./Cons. | 73.6 | 42.9 | 47.4 | 66.5 | 57.6 |
| Mean | 72.4 | 54.8 | 56.8 | 68.2 | 63.0 |

Table 10

Mean Latency (Sec.) of Correct Responses on Inference Post-Test, Classified By Correct Response

| Group | Test Sentence Type | | | | |
|---|---|---|---|---|---|
| | D-Yes | D-No | R-Yes | R-No | Mean |
| Antec. | 12.3 | 12.2 | 13.3 | 11.0 | 12.2 |
| Conseq. | 11.9 | 16.8 | 14.7 | 12.8 | 14.1 |
| Ant./Cons. | 12.3 | 12.7 | 14.2 | 11.9 | 12.7 |
| Mean | 12.2 | 13.9 | 14.1 | 11.9 | 13.0 |

was statistically significant whereas neither of the two main effects was. Before discussing the implications of this finding, the details of the statistical analysis will be reported.

Recall that only the reaction times associated with correct responses were included in the analysis, which produced an unequal number of observations per cell. Hence multiple regression had to be used to partition the variance. The partitioning was accomplished by what is sometimes called a (classical) experimental design analysis, a generalization of what Overall and Spiegel (1969) call "Method 2". In this method the component sums of squares need not add up to the total sums of squares (and indeed do not in the ANOVA tables of this report). For completeness, it should also be noted that analyses of repeated measures designs as large as the present one can be accomplished on most computers with the usual statistical packages only by recourse to special coding techniques (Pedhazur, 1977).

Extremely long reaction times (i.e., those over 3 standard deviation units from the cell means) were excluded from analysis. These comprised 1.9% of all correct responses. Nine subjects were excluded from analysis because they had very few correct responses. The rationale for this exclusion is that these subjects scored so low as to suggest they were answering the verification questions backwards. For, if the participants were responding randomly to the "Yes" or "No" questions, then the total number of correct responses would be a binomial random variable with parameters .5 and 24. Then the 95 percent confidence interval for the percent correct would be approximately 30% to 70%. It is therefore a safe bet that the nine participants scoring less than 30% were not responding randomly. But neither were they responding in the proper direction. (Parenthetically, a methodological note is worth inserting. The fact that only 34 of the 63 participants scored above 70% correct attests to the difficulty of the task. Perhaps this difficulty is attributable in part to the unfamiliarity of the verification task itself rather than to an unawareness of causal relations. Several participants commented that they became confused during the test about whether or not they were "doing it right" and may have answered some of the questions backwards. If this verification task is pursued in future research, feedback of correctness given after each response and a slower paced test -- there was essentially no intertrial period in the procedure of the present study -- might help reduce this confusion and increase performance.)

As noted, there are two alternative ways of classifying the latency data for statistical treatment. The one way was illustrated by Table 8, where the three independent variables were group, directness, and relation (antecedent or consequent, abbreviated A or C, respectively, in the table). The other classification appears in Table 10, where "relation" is replaced by "response" ("yes" or "no") as the third independent variable. Two ANOVAs, one for the Table 8 classification and the other corresponding to Table 10, were calculated. Results are summarized in Tables 11 and 12, respectively.

In both tables, both the Groups and the Subjects main effects are highly significant, while the Items effect approaches significance. Further, in Table 11 the Relation effect approaches significance (though in Table 12 the Response main effect does not). What the borderline Relation effect suggests is that consequent test alternatives are more rapidly responded to correctly than are antecedent alternatives (note the column means in Table 8). At the same time, the Group effect showed that correctly deciding whether a test alternative was an antecedent of the key sentence was faster than correctly deciding whether an alternative was a consequent. Put another way, of the four combinations of Group and Relation, the case of "Was test sentence an antecedent of key sentence?" combined with a consequent test sentence led to the fastest correct responses; the case of "Was test sentence a consequent of key sentence?" combined with an antecedent test sentence led to the slowest correct responses.

The marginal effect of Relation, however, is overshadowed by the significant interaction of Directness with Response (Table 12), the X-interaction already noted. The interaction holds over all three groups. Recall from Table 9 that the direction of the effect is one of faster correct "yes" responses than "no" responses to direct test alternatives, but faster correct "no"s than "yes"s to remote alternatives.

Originally the intention was that, if such an interaction should be found, we would then investigate text memory by seeing how this X-interaction varied under different levels of processing. The "depth of processing" was to be varied by varying the task preceding the inference test (e.g. from casually reading the text to studying it). It was thought this would be a profitable venture due to the correspondence of this obtained interaction and the "semantic-relatedness" effect found in semantic memory research using sentence verification. The "semantic-relatedness" effect is an

Table 11

ANOVA with "Relation" as one Independent Variable

| Source | SS | df | MS | F | P |
|---|---|---|---|---|---|
| Total | 53012.0 | 880 | | | |
| Groups | 616.0 | 2 | 308.0 | 7.70 | p<.001 |
| Subjects | 8753.9 | 20 | 437.7 | 10.94 | p<.001 |
| Residual | 27363.7 | 684 | 40.0 | | |
| Items | 893.3 | 5 | 178.7 | 2.01 | .05<p<.10 |
| SxI | 8897.5 | 100 | 89.0 | | |
| Relation | 453.3 | 1 | 453.3 | 3.24 | .05<p<.10 |
| GxR | 19.6 | 2 | 9.8 | .07 | |
| SxR | 2802.7 | 20 | 140.1 | | |
| Directness | 5.3 | 1 | 5.3 | .06 | |
| GxD | .5 | 2 | .3 | .00 | |
| SxD | 1685.8 | 20 | 84.3 | | |
| RxD | 46.7 | 1 | 46.7 | .67 | |
| GxRxD | 72.6 | 2 | 36.3 | .52 | |
| SxRxD | 1388.4 | 20 | 69.4 | | |

Table 12

ANOVA with "Response" as one Independent Variable

| Source | SS | df | MS | F | P |
|---|---|---|---|---|---|
| Total | 53012.0 | 880 | | | |
| Groups | 572.0 | 2 | 286.0 | 7.15 | p<.001 |
| Subjects | 8696.6 | 20 | 434.8 | 10.87 | p<.001 |
| Residual | 27381.7 | 684 | 40.0 | | |
| Items | 857.7 | 5 | 171.5 | 1.92 | .05<p<.10 |
| SxI | 8942.6 | 100 | 89.4 | | |
| Response | 22.8 | 1 | 22.8 | .17 | |
| GxR | 121.9 | 2 | 61.0 | .45 | |
| SxR | 2695.7 | 20 | 134.8 | | |
| Directness | 6.4 | 1 | 6.4 | .08 | |
| GxD | 1.6 | 2 | .8 | .01 | |
| SxD | 1639.7 | 20 | 82.0 | | |
| RxD | 405.0 | 1 | 405.0 | 5.92 | .01<p<.025 |
| GxRxD | 8.0 | 2 | 4.0 | .06 | |
| SxRxD | 1367.2 | 20 | 68.4 | | |

interaction between the degree of association between the subject and predicate of the sentence to be verified (corresponding in the present study to the Directness of the test sentence to the key sentence) and the truth value of the sentence (corresponding in the present study to the Response ("Yes-true" vs. "No-false")). In semantic memory studies a typical finding is that for true ("Yes") sentences correct verification is quicker if the subject and predicate are high related (direct) than if they are less related (remote), e.g. A robin is a bird vs. A penguin is a bird, respectively. The other half of the "semantic relatedness" effect is the typical finding that with false ("No") sentences correct verification is slower if the subject and predicate are high related (direct) than if they are less related (remote), e.g. A robin is a mammal vs. A robin is a car, respectively. The original strategy was to take advantage of the fact that text memory can easily be manipulated (whereas its difficult to manipulate semantic memory) to help elucidate the nature of this relatedness effect as well as explore the structure of text memory. However, recent publications on verification time and semantic memory (Corbett & Wickelgren, 1978, also McClosky & Glucksberg, 1979) have cast doubt on the utility of using verification times to infer the structure of memory. The conclusions of these recent studies is that many typical findings in sentence verification studies depend on decision processes rather than on memory structure.

Therefore it seems that our reaction time results cannot take us much closer to our original goal of explaining text memory. Nonetheless, they are important, because the scope of verification theories must be broadened to account for this interaction of response and directness occuring even in the verification of causal relations. Offhand it seems the processes underlying verification of causal relations would be a good deal more complex than the verification of the class-inclusion relations usually employed in verification studies. But perhaps the same processes are used, and it just takes longer to reach a decision. At any rate, the original research strategy described above might be modified, in light of the recent publications, to yield predictions that distinguish among different verification theories.

Experiment 13.

In this study and the next, the use of a sentence sequencing task to examine comprehension of text coherence relations was resumed from Experiment 9. Recall that the most notable finding from that pilot work was

that a participant's performance on the inferencing post-test was more highly correlated with the agreement of his own sorting order to a theoretically ideal premise ordering than to an ideal presupposition ordering.

Further research with this task was deferred until the computer software to enable control the experiment had been developed. An advantage of using a computer terminal rather than a manual card-sorting is that it allows detailed observation of the sequencing task, beyond just the resulting final order of the sentences. Namely, the steps taken to arrange the sentences and the time between steps can easily be recorded. It is hoped that this "time between steps" will provide an index of the subjective cohesion between sentences.

Method. Again, the experimental text was a version of Circle Island. First the participant was given all of the sentences listed in random order on a sheet of paper. He was instructed to read through the list once. Each sentence was prefixed with an identifying alphabet letter, assigned randomly without replacement. The participant was to arrange the sentences in a coherent textual sequence, using the sentence identifier letters to "tell" the computer his sentence "moves". For example, the first three commands a participant types into the computer terminal might be "put Y in 1", "put N after Y", and "put E before N". After each command, the text is built on the terminal screen in the order specified by the participant. Because the text is too lengthy to fit all at once on the screen, there are two special commands, "Up" and "Down", which allow the participant to position and study, at will, the text he is building. After each command, that command and the time since the previous command are recorded. The participant is shown how to use the terminals and given two practice texts before rearranging the experimental passage. Thus far, out of 19 participants none has had trouble in learning how to use the commands or terminals. There is, however, much variation in the time taken to complete the experiment; the range was from 45 to 90 minutes. Because this variation is almost certainly also present in the time taken to rearrange the experimental passage, a normalization of the time-between-commands to the proportion of time-between-commands to total time will be made before other analyses are carried out. (Though the real time data may be of interest in later analyses.)

Results. Data analyses are still in progress, not yet in reportable form. Time was devoted to writing a computer program to execute the analyses.

The program will plot the results of an INDSCAL analysis on the Computing Center's CALCOMP plotter. Since the results from INDSCAL are inherently spatial, those results are almost always plotted for presentation and analysis. And since there could be a great many INDSCAL analyses during the course of this project (resulting in perhaps as many as 50 two-dimensional plots) this automated plotting procedure will quicken later analyses. Here the aim of the INDSCAL application is to arrive at an empirically "ideal" ordering based on the data. If this analysis is successful, the representation itself will be of much interest. Further, it may also allow us to characterize the empirical orderings not only in terms of the amount of deviation from an ideal but also in terms of the type of deviation.

Experiment 14.

A closely related study still in progress uses the sentence sequencing task to investigate individual differences in subjective text cohesion. As suggested earlier, one possible difference is that some readers may prefer to have consequents be near their antecedents, while others may prefer that consequents be near their (topical) presuppositions. For some texts, including Circle Island, these two orderings turn out to be nontrivially different.

Method. Except for the addition now of post-tests, the procedure was identical to that of the previous experiment. Ten participants received the same post-test as had been used in the adjunct connectives work (project IV). It was expected that readers inclined to a premise organization in the sequencing task would do relatively well on this particular post-test. Another ten participants were asked simply to write an outline of the passage they had read. Now the expectation was that readers who favor a presupposition organization would do relatively well.

Results. Data analyses are at the same stage as for the preceding study. An aim of the INDSCAL analysis this time will be to see if a presupposition dimension and a premise dimension emerge in the sequencing data. If so, then the ordering of the subjects' points along these two dimensions should reflect the relative importance that the different subjects place on these dimensions. Correlations will then be calculated between the rankings on these dimensions and the rankings of performance on the two post-tests.

VI . Toward representing a text as a "proof" in Natural language logic. (Edward Crothers)

Under this heading a side theoretical venture will be summarized. The undertaking would appear to be of value for natural language logic, but upon investigation it turned out to be perhaps more a formalization of existing representations than a fundamentally new theoretical direction. The "proof" formalization was not entirely without application, however. It did lead to the Nobel Prize revision used in Experiments 4 and 5 of the text revision project.

Besides the abstract interest of a proof-like representation of a text, the method is potentially appealing in another respect. It can be claimed that the derived quasi-proof is in fact a canonical representation of the text in question (reminiscent of a "kernel sentence" in early transformational grammars), as contrasted with a representation that is arbitrary in important respects. Hence, a fairly nonarbitrary way of comparing two or more texts is on the basis of their canonical forms. For simple narratives a story grammar is a canonical form, though as argued previously herein the particular grammars proposed to date are descriptively inadequate. The proof notion is weaker descriptively and not even a generative grammar at all. On the other hand, it is applicable to all discourse types. Naturally, argumentative texts would be expected to exhibit a sentence sequence closer to the proof form than would, say, narrative texts.

First a caution is in order as to the present status of my work on the proof analogy. As of now the formulation is best regarded as a technical tour de force whose broader theoretical and experimental significance remain to be established -- a method in search of a theory, in search of an application. This state of affairs is not too unlike that of having formulated a new form of sentence grammar without yet knowing its theoretical implications or empirical worth. Also, an unattractive feature of the method as now conceived is that it is tedious to apply to text passages beyond a few paragraphs in length.

The method itself can be summarized informally as follows. A text is viewed as a "proof" in the sense that its topic sentence is regarded as a consequent "theorem" derived from the sentences that immediately develop it, they in turn from their own antecedent expansions, and so on. The "proof", sketchy though it

may be, can terminate when all of the explicit text sentences have been included. Or, if a representation of inferences is desired, the proof can be expanded by including just those inferences that are recognized by the principles of inferential text coherence. Or, purely pragmatically, the "proof" can terminate even before all of the explicit text details have been included, by resorting to the familiar admonition "Details are left to the reader."

That the representation is only a quasi-proof is obvious. What is interesting is the sense in which it is a proof and the sense in which it falls short. In both significant respects, namely the axioms and the rules of derivation, the representation is at best only an analogue to a proof within a logical calculus. The "axioms" of a text's representation are taken to be those propositions that (i) are present in that text (again, either explicitly or implicitly, depending on how complete a "proof" is sought), and (ii) are not consequents of other propositions within the text. In other words, the first significant limitation on the proof character is that the "axioms" are specific to the text in question instead of being some minimal group of fundamental premises implying many theorems. The other limitation, predictably enough, is that the derivation of one proposition from another is much looser than in true deductive systems. This limitation is well known and has been discussed many times, though in the context of proposition-to-proposition implication rather than treating a text as a unified proof. The familiar distinction between logical and semantic consequence (e.g., van Dijk, 1973) reflects this tighter-looser distinction.

The method will be illustrated here with the first six sentences of the two-page Nobel Prize article.

The 1976 Nobel Prize in Medicine

S1. The 1976 Nobel prize in medicine was shared by two physicians, Baruch S. Blumberg of the Institute for Cancer Research in Philadelphia and D. Carleton Gajdusek of the National Institute for Neurological Diseases and Stroke. S2. Blumberg and Gajdusek were each honored for his discovery of the origin and mechanisms of certain infectious diseases.

S3. Blumberg discovered the origin of one kind of hepatitis, and perhaps one kind of cancer as well. S4. His work began in the early 1960s. S5. Originally he was trying to discover how genetic variations between human populations influence susceptibility to hepatitis and other diseases. S6. Therefore he studied some 100,000 blood samples taken from genetically different populations.

The derived "proof" appears in Figure 1. "Proof" lines corresponding to explicit sentences are marked by asterisks. It can be seen at once that the "proof" rapidly becomes very lengthy, owing to the many inferences (here, five times the number of explicit sentences) required to formalize it. This is true because the additional inference principles for proof-like cohesion yield even for simple texts many more inferences than those derived from the principles already established in the theory of inferable text cohesion as applied in the preceding section of this proposal. Nonetheless, this Nobel Prize excerpt does offer a simple introduction to the method, because it will be seen that only one transformation from canonical sequence to surface sequence is required.

To simplify matters we take the sentence rather than the constituent proposition as the unit of analysis, going to a proposition level analysis only when the quasi-proof requires. We proceed sentence by sentence. It is more efficient to proceed from consequent sentence to antecedent sentence, which for texts in general need hardly be mirrored by the sentence sequence. However, to facilitate the exposition here the excerpt selected is one whose sentence sequence rarely departs from a consequent-to-antecedent order. First, the inference that S1 is a (semantic) consequent of S2 can be written as

1. To prove: S1
2. Proof: S2 & S2 $\longrightarrow$ S1
:
Proven: S2 & S2 $\longrightarrow$ S1
. Proven: S1
:

What has been done here is proof reduction: decomposing the proof of S1 into that of the conjunction of S2 with S2 $\longrightarrow$ S1. In other words, in the canonical form the

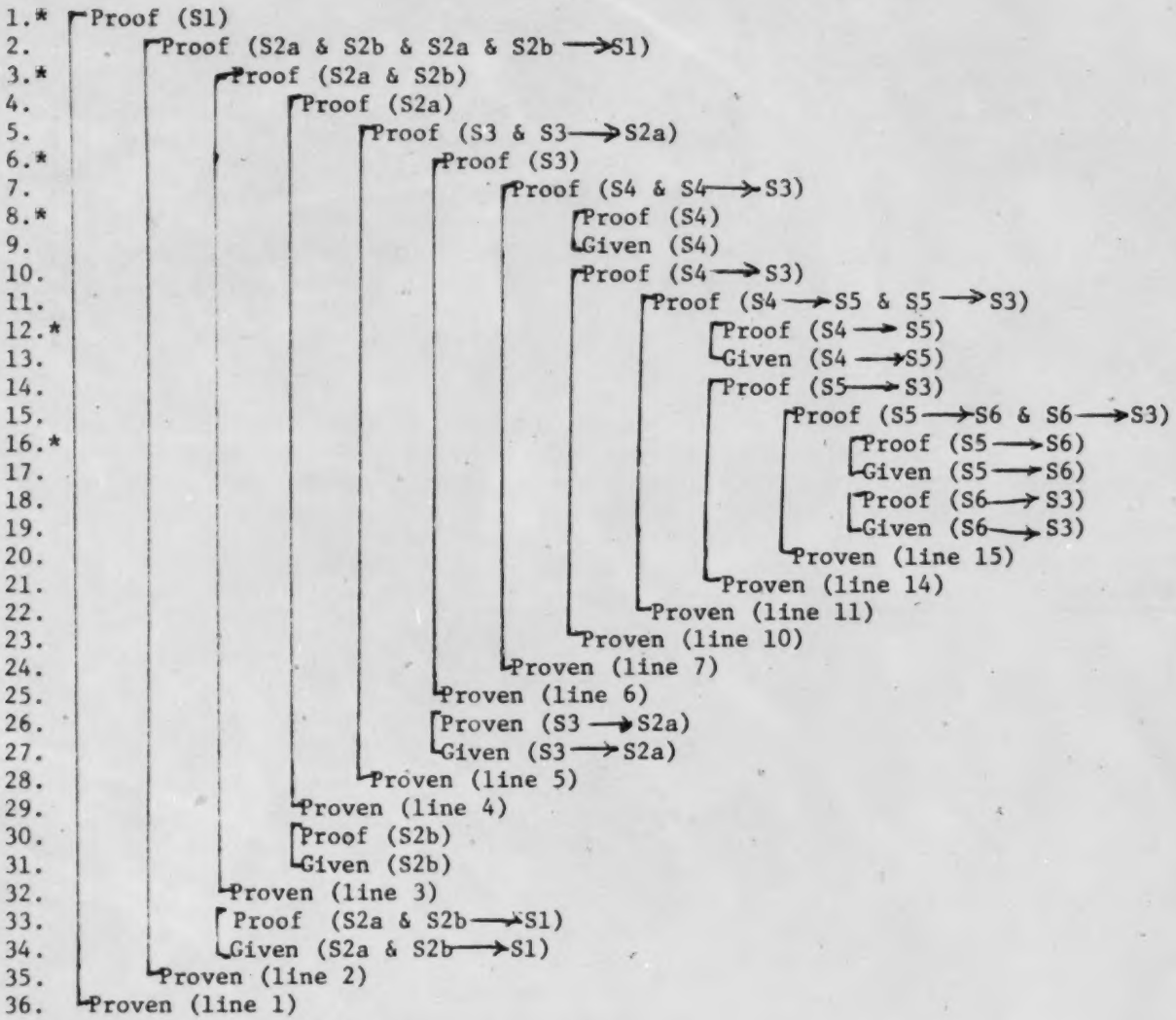

1.* Proof (S1)
2. Proof (S2a & S2b & S2a & S2b → S1)
3.* Proof (S2a & S2b)
4. Proof (S2a)
5. Proof (S3 & S3 → S2a)
6.* Proof (S3)
7. Proof (S4 & S4 → S3)
8.* Proof (S4)
9. Given (S4)
10. Proof (S4 → S3)
11. Proof (S4 → S5 & S5 → S3)
12.* Proof (S4 → S5)
13. Given (S4 → S5)
14. Proof (S5 → S3)
15. Proof (S5 → S6 & S6 → S3)
16.* Proof (S5 → S6)
17. Given (S5 → S6)
18. Proof (S6 → S3)
19. Given (S6 → S3)
20. Proven (line 15)
21. Proven (line 14)
22. Proven (line 11)
23. Proven (line 10)
24. Proven (line 7)
25. Proven (line 6)
26. Proven (S3 → S2a)
27. Given (S3 → S2a)
28. Proven (line 5)
29. Proven (line 4)
30. Proof (S2b)
31. Given (S2b)
32. Proven (line 3)
33. Proof (S2a & S2b → S1)
34. Given (S2a & S2b → S1)
35. Proven (line 2)
36. Proven (line 1)

Figure 1. "Proof" of the excerpt from the Nobel Prize text. The proof of S3, lines 7-24, is a transformation of the canonical form.

proof works backwards from consequents to antecedents, to their antecedents in turn, and so on. Once the line 2 conjunction has been established, S1 will follow by modus ponens, a standard axiom of a natural logic. The two "proven" counterparts are inserted as a formality. But they are left un-numbered for the time being, until the corresponding proofs have been completed. Because every line is a candidate for expansion, line 1 is no different, and the locution "To Prove" there can be replaced by "Proof" as in the other lines.

The next step is to establish line 2. However, the third sentence, namely

> S3. Blumberg discovered the origin of one kind of hepatitis, and perhaps one kind of cancer as well,

is seen to furnish a reason for only the first conjunct, not the second, of line 2. Accordingly, the proof is decomposed into

```
1.*  Proof:  S1
2.     Proof:  S2 & S2 ⟶ S1
3.*      Proof:  S2
:          :
:          :
:        Proven:  S2
:        Proof:  S2 ⟶ S1
:          :
:          :
:        Proven:  S2    S1
:      Proven:  S2 & S2 ⟶ S1
:    Proven:  S1
```

Here obviously the axiom is conjunction distribution: Proof(Si & Sj) is decomposable into Proof(Si) & Proof (Sj). Referring again to S3, a further preparatory conjunction distribution is also needed, because S3 provides a reason for only the Blumberg conjunct, not the Gajdusek conjunct, of S2. Numbering these conjuncts S2a and S2b respectively, the proof can continue by rewriting lines 2ff. (or, alternatively, by recording the distribution at line 4):

```
1.*  Proof:  S1
2.     Proof:  S2a & S2b & S2a & S2b → S1
3.*      Proof:  S2a & S2b
4.         Proof:  S2a
5.           Proof:  S3 & S3→ S2a
6.*            Proof:  S3
:
:
```

Continuing, the next two sentences are:

> S4. His work began in the early 1960's. S5 Originally he was trying to discover how genetic variations between human populations influence susceptibility to hepatitis and other diseases.

Two sentences are considered together here, to make the point that depending on the inferred structure it is often easier to write the subproofs separately and then embed them in the overall proof, instead of proceeding sentence by sentence. But regardless of the tactic (and even regardless of whether a proof is being compiled at all), at some point it becomes necessary to infer the connection between S4 and S5. Auxiliary syntactic or paraphrastic operations are often helpful; here S5 can be restated by recovering the antecedents he for Blumberg and in the early 1960's for Originally. Now a semantic judgment, outside the theory, is needed in order to infer the connection between S4 and S5. Let us take S4 as an antecedent, rather than a consequent, of S5.[1]

---

[1] The interpretation of S5 is possibly ambiguous, leading to an ambiguity as to the connection to S4. On one reading, trying in S5 expresses a purpose or motive, hence the meaning is not unlike that of Originally, he wanted ···. On this view S5 is an antecedent of S4, because motives are antecedents for actions. However this view encounters difficulty with S6; after S5 → S4 now and S5 → S6 next, the connection between S4 and S6 is unclear. An alternative reading, the one adopted here, is that S4 is actually a presupposition for S5. That is, S5 could be restated roughly as Originally he began working to discover···. Because in the theory presuppositions are treated as antecedents rather than as consequents, on this preferred view the inferred relation is reversed; S4 is an antecedent of S5. The moral is that a representation, be it a graph, a "proof", or whatnot, is just that -- a representation. It represents the inferences drawn, but does not replace one's semantic intuitions by which those inferences were drawn in the first place.

Continuing, the last sentence of the excerpt is:

S6. Therefore he studied some 100,000 blood samples taken from genetically different populations.

Clearly, this is a consequent of S5. The last three sentences, then, connect as S4 $\rightarrow$ S5 $\rightarrow$ S6. Referring back to S3, each of these three is understood as an antecedent of it. Hence the canonical form would embed S3 first, then S6, then S5, and finally S4. But the actual text sequence is a transformation of that sequence, the earlier links in the chain being stated first. Hence in conjunctions of the form Si & Si $\rightarrow$ Sj, it is now the second conjunct that gets expanded:

```
6.* Proof:  S3
 7.    Proof: S4 & S4 → S3
 8.*     Proof: S4
 9.      Given: S4
10.      Proof: S4 → S3
11.        Proof: S4 → S5 & S5 → S3
12.*         Proof: S4 → S5
13.          Given: S4 → S5
14.          Proof: S5 → S3
15.            Proof: S5 → S6 & S6 → S3
                        etc.
```

Finally, the subproofs are routinely terminated, from more deeply to less deeply embedded, thus completing the figure. Observe that because we have here excerpted only the beginning of the passage, when the deferred conjuncts such as S2b for Gajdusek eventually come up to be proven they are simply noted as given. Had the analysis as derived for the full two pages of Nobel Prize been presented here, many such conjuncts would have been proven rather than left as given.

To digress briefly, an interesting conjecture considering the text's sequence of all six sentences is that the observed pattern, namely S1 $\leftarrow$ S2 $\leftarrow$ S3 $\leftarrow$ S6 $\leftarrow$ S5 $\leftarrow$ S4, is probably not at all unusual. First a thematic conclusion is set forth, accompanied by a few direct subthematic antecedents of it. Then the text goes back to a causal, chronological, or logical beginning and proceeds towards those major themes already enunciated.

This example conveys the idea of the method. The notion of a canonical proof is well defined: it proceeds backwards from consequents to antecedents recursively. As presently conceived an element of

arbitrariness does remain, namely the expansion of unconditional conjuncts Si before conditional ones Si $\rightarrow$ Sj. Whether or not this can be justified remains to be seen. Beyond the canonical form, out of indefinitely many possible proofs the method indeed selects one that represents the surface text. Depending on the latter, conjuncts are or are not separated, subproofs are or are not expanded, and transformations from the consequent-to-antecedent order are or are not introduced. One important variation found in some texts, though not in the above example, is topicalization of antecedents. For example, a cause may be topicalized by appearing in a title or lead sentence, followed by introducing one or more effects. Within the proof framework this is conceived as antecedent-raising, producing (b) from (a), where Si denotes a cause and Sj, Sk a pair of effects:

| | |
|---|---|
| Sj & Sk | Si |
| Si & Si $\rightarrow$ Sj & Sk | Si $\rightarrow$ Sj & Sk |
| Si | Si & Si $\rightarrow$ Sj & Sk |
| Si $\rightarrow$ Sj & Sk | Sj & Sk |
| (a) | (b) |

Other familiar surface text phenomena are easily represented. In addition to the prevailing pattern in the excerpt, namely "Proof" lines explicit and "Proven" lines implicit, the other two possibilities appear frequently in texts. Conclusions may be stated only after their arguments. Or they may be advanced at the outset and reiterated at the closing. The asterisk notation would indicate accordingly. Quite different text properties are representable as well. Because it is the surface text and not one's own knowledge as to the evidence for a particular proposition that is being represented, even if a text asserted logically impossible or semantically implausible connections, they would remain as such in the proof. Indeed, each exact violation would be more precisely identified.

What are some of the advantages and disadvantages of representing a text as a proof compared to representing it as a graph? In the first place, certainly a graph can be derived from a canonical form, as is done with a story grammar or sentence grammar. The same could be done with the form proposed here; there is no objection to graphs in principle. It remains to be seen which of the two modes, proof or graph, is the more perspicuous. One possible reason for favoring the

proof format should not be overlooked, however. The description can be conveniently annotated, as in a conventional proof, by showing on each proof line the particular natural logic rule invoked. In general, connections of the form Si $\rightarrow$ Sj can exemplify any of several types of rules, notably logical axioms, lexical meaning postulates, and semantically possible or necessary implications, not to mention explicit assertions. Another possibly interesting analogy to a true proof is that some texts, such as Circle Island, contain propositions that might be viewed as "theorems." These, graphically identifiable by their multiple and remote connections, are major intermediate results -- major in that they later enter as antecedents for not one but several consequents.

For texts in general, most published graph representations have not been based on grammars. Here the comparison to the proposed approach deserves comment. As noted, a serious drawback to the proof idea is that, predictably, its explicitness renders it lengthy, tedious to derive. This problem need not be insuperable, because in practice one can choose to derive only the intricately interconnected segments of a passage. As compensation for the labor, the outcome is sometimes illuminating. Graph methods of text representation currently in vogue in psycholinguistic research tend to obscure the proof-like quality, because their descriptions are not very abstract. Antecedence-consequence relations are not abstracted from the particular connectives, and there is no description in terms of a standard form and a departure therefrom. For one thing, this makes the identification of implicit propositions more difficult. Having a canonical form assists in knowing what to look for, which aids not only in finding it but also in knowing when it isn't there. Another benefit from a canonical form is that if logically or semantically related topics are presented dispersed in a text, either because of stylistic license or because of disorganization, the proof method is suited to describing the adopted organization in terms of a standard and a variation thereon.

By the same token, the method is applicable to comparing multiple texts. If a comparison is quite direct, a sophisticated method would be superfluous. For example, the texts might differ only in what is explicit vs. what is implicit. But in other instances the comparison might be more complex. There

might be materially different sequencings and parsings of the same content, as already discussed. Other possible relations between texts are also suggested by the proof framework. However, at this time I have not studied them seriously, hence can list only a haphazard mixture of possibilities:

1. One text might be included in another, in that the latter expands the former.

2. Two texts might overlap, in that each expands something left as given in the other.

3. Two texts might have identical outer subproofs but divergent inner, more embedded details.

a. They might be identical only at some more abstract level beyond the proof itself. That is, they might be alternative instantiations of the same script.

b. Less abstractly, they might have identical themes but rather unrelated expansions.

4. Conversely, they might have identical antecedents but diverge at higher levels to different conclusions or outcomes.

a. As one special case, they might have mutually exclusive consequents, owing to the inherent noncertainty of natural inference.

b. They might have mutually compatible consequents, of which the one implies the other. In this case the one text is "stronger" than the other. From the same antecedents it has developed more consequents. (Compare possibility number 1 above.)

5. Purely structurally, they might have the same antecedence-consequence development but totally different content.

## VII. Representing parallelism in texts: Circle Island (Edward Crothers).

The report here of this project will be fairly brief, because all except the final stages of the work have already been described elsewhere (Crothers, 1978, 1979b, Chapter 8). Some space will be devoted to the further development of parallelism representation beyond those publications.

Rather surprisingly, the work actually grew out of an original intention of a quite different sort, namely to extend the "proof" analysis to a second passage. Circle Island was selected because it offered several advantages: it is a narrative, hence presumably challenging to a proof-based approach; it is notorious for its many inferences, again a challenge; and the availability of past analyses by others provided a perspective from which to evaluate any new representation. Upon undertaking the analysis it was soon realized that Circle Island exemplified an intricate, though actually not unusual, pattern of text development. The difference from Nobel Prize went deeper than the sheer number of inferences. What I have called "parallelism" was prominent in Circle Island. Once this was recognized, the proper working out of the detailed representation became the goal of the theoretical undertaking. The further goal of formalizing the representation as a proof had to be postponed until the graph representation itself had been derived. The latter proved to be an instructive task of not inconsiderable magnitude, and consequently the proof formalization project was never resumed.

The earlier stages of the project, cited above, proceeded by first appealing to one's understanding of the passage to establish what it is that must be represented, and then conceiving a convenient form of representation. The outcome of the analysis to that point was a representation of Circle Island that (i) is a beginning toward handling parallelism in the story, and (ii) views the story as an instantiation of a story grammer (though without proposing any grammar for stories in general). The more recent, unpublished development is first an advance in great detail on point (i), the treatment of parallelism. But the payoff is that when this "microstructure" has been set forth in detail the result is an appreciably deeper understanding of the "macrostructure", the overall organization among major constituents of the grammer of this particular narrative. Reporting of details here will be limited to an example of parallelism and a sketch of the final derived story grammar macrostructure.

First, the narrative version used here is:

Circle Island

1. Circle Island has good soil. 2. The island has few rivers and hence a shortage of fresh water.

3. The main occupations on the island are farming and cattle ranching. 4. The farms on the island are affected by the lack of fresh water. 5. The ranches are less affected by the lack of water. 6. The ranchers are more prosperous than the farmers. 7. No ranchers are farmers.

8. The island is run democratically. 9. Each issue is brought to a vote. 10. The majority will on each issue is decided by the majority vote. 11. The actual governing body is a senate, whose job is to carry out the will of the majority. 12. The senate consists of the richest men, because it is felt that the richest men are the best proven administrators. 13. All senators are ranchers.

14. The farmers wanted the government to build a canal across the island. 15. An island scientist had discovered a way of converting salt water to fresh water. 16. Therefore if a canal were built the farmers could convert sea water brought by the canal into fresh water to use to cultivate the island's central region.

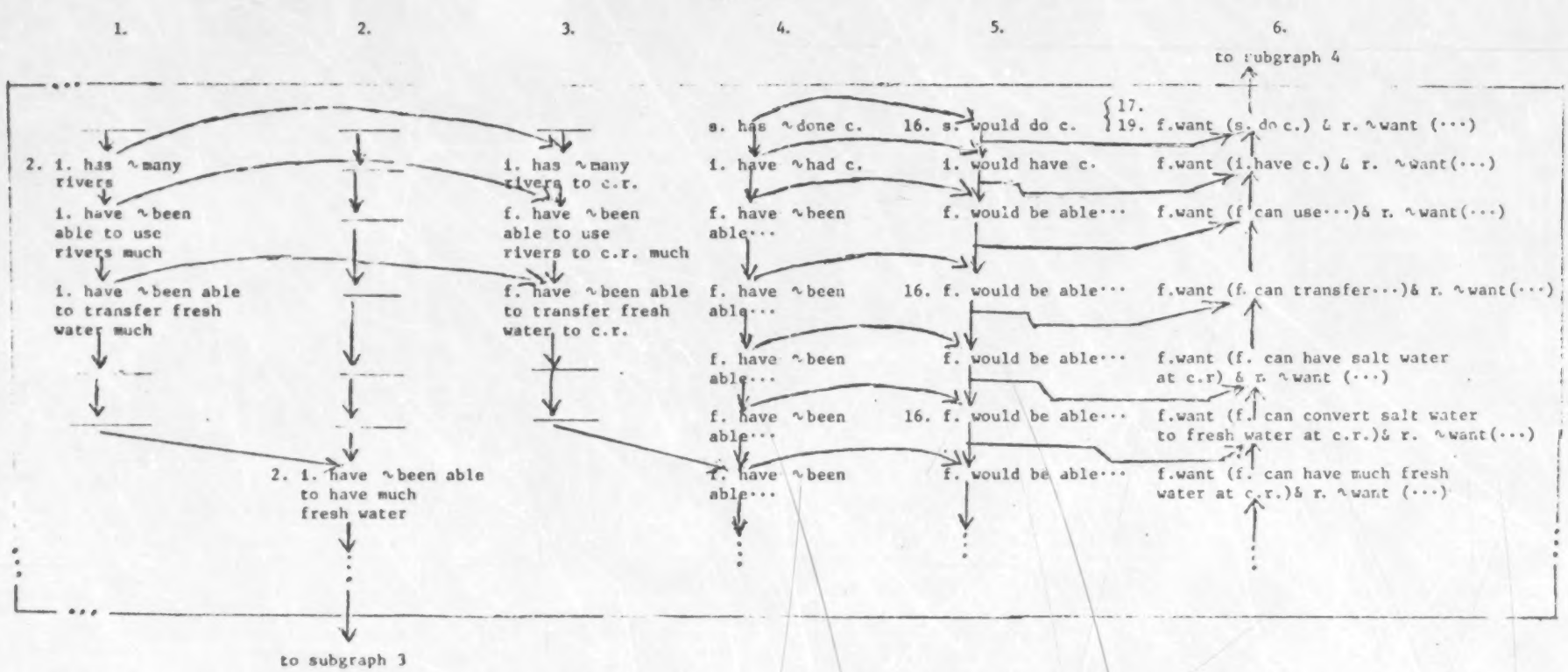
17. The ranchers did not want the government to build a canal across the island. 18. They did not want the farmers to become able to cultivate the central region. 19. If a canal were to be built at all, the ranchers opposed a large canal. 20. A large canal would disturb the island's ecology.

21. The idea of building a canal turned into an issue. 22. The idea of building a canal was brought to a vote. 23. All the islanders voted. 24. All the farmers voted in favor. 25. The ranchers all voted against it. 26. The vote of the majority was in favor of building a canal.

27. The senate consented to begin a project to build a canal across the island. 28. But the senate allowed only a small canal trench to be dug. 29. The canal trench was so shallow that no water would flow into it. 30. The canal project was abandoned.

31. Many of the islanders were extremely angry. 32. It was likely that there would be a civil war.

For illustrative purposes here consider only those segments of the story that develop the origin of conflict between the farmers and the ranchers; these segments comprise most of paragraphs 2, 4, and 5 of the story. Figure 2 shows the representation. The figure is to be read as follows. For brevity, only the upper right-hand quadrant of the subgraph is displayed. Propositions that correspond to sentences


Chain No.
1.
2.
3.
4.
5.
6.
to subgraph 4
2. i. has ∿many rivers
i. have ∿been able to use rivers much
i. have ∿been able to transfer fresh water much
2. i. have ∿been able to have much fresh water
to subgraph 3
i. has ∿many rivers to c.r.
f. have ∿been able to use rivers to c.r. much
f. have ∿been able to transfer fresh water to c.r.
s. has ∿done c.
16. s. would do c.
17.
19. f.want (s. do c.) & r. ∿want (···)
i. have ∿had c.
i. would have c.
f.want (i. have c.) & r. ∿want(···)
f. have ∿been able···
f. would be able···
f.want (f. can use···)& r. ∿want(···)
16. f. would be able···
f.want (f. can transfer···)& r. ∿want(···)
f.want (f. can have salt water at c.r) & r. ∿want (···)
f.want (f. can convert salt water to fresh water at c.r.)& r. ∿want(···)
f.want (f. can have much fresh water at c.r.)& r. ∿want (···)

-70-

Figure 2. Upper right hand section of subgraph 2, the geo-economic subgraph. Notation: Ellipses in the fourth and fifth columns refer to the sixth column. Ellipses in the second conjuncts of the latter refer to their respective first conjuncts. Abbreviations: i.= island(ers), f.= farmers, r.= ranchers, c.r.= central region, s.= senate, c.= canal. Symbols: ∿ = negation, — = null node. Numerals refer to corresponding sentences of surface text.

77

BEST AVAILABLE COPY

76

of the quoted text are numbered accordingly. As anticipated, most of the propositions and connections are inferred. Arcs in the graph are directed from antecedent to consequent. (Note that in the graph there can be higher-order relations, where one or both arguments is itself a subgraph of more than one node. Examples occur in the last two columns, where propositions of the form If X then Y combine with ones of the form Want Y to yield consequent Want X). Above all, the figure is a series of proposition chains paralleling one another. The first four chains or columns represent the actual situation, the fifth the conditional situation, and the sixth the desired (or not desired) situation. Further, the actual situation can be thought of as itself divided into a situation on the island as a whole (chains 1 and 2) and a restriction to the central region (chains 3 and 4). Hence one parallel is between the first two chains and the second two. Another parallel, between the actual and desired situations in the central region, comprises the second parallel within each of these pairs of chains. That between the third and fourth chains, for instance, turns on rivers and canals as alternate means of supplying water, a fact hardly immaterial to one's understanding of the story.

Since most of the parallels are only inferred for Circle Island, it is pertinent to consider both the question of general criteria for inferring parallels and the question of how they are actually inferred. A reasonable criterion for a text level approach, and one that would pursue the analysis to an intermediate depth, is to include only those chains that either contain one or more propositions corresponding to explicit sentences or parallel chains satisfying the first condition. That is, implicit chains parallel only to other implicit chains would be excluded. By this criterion each of the six chains shown in the figure does indeed belong in the representation.

The figure also suggest observations on how parallels are actually inferred. As with other kinds of inference, semantic knowledge must of course develop the textual testimony. This point might be less apparent for parallels, because a glance at the figure perhaps suggests that successive chains have been generated from one another by a series of mechanical substitutions. However, this is not the case in general. For instance, while it is true

that some of the entries in the third column are merely the restriction to central region of their counterparts in the first column, nonetheless others also change the quantity term from a relative (much, many) to an absolutive (not at all). Another observation is that the arcs do not necessarily reflect the sequence in which the inferences are drawn, any more than they reflect the sequence of surface sentences. A routine example is that presuppositions, such as those in chain 4, are in fact often never asserted as such, being inferable only retrospectively from their presupposing propositions, such as those in chain 5. Sometimes either direction of inferencing, namely forward or backward, is plausible. For example, the third chain on the link between few rivers and inadequate water might be inferred at once from one's knowledge that the distribution of water, and not merely the distributable supply, affects the available quantity. On the other hand, this mundane but relevant inference about rivers might pass unnoticed until its parallel about canals is encountered. Finally, when extending a parallel beyond its explicit "anchors", a reader is not wont to stretch it to its imaginable limit. An inference that the farmers wanted seats on the senate, though not utterly implausible, does take liberties with the text.

Figure 3 depicts the global structure ultimately inferred. The nodes are named with proposed story grammar constituents (of course, only those constituents found in this particular story). By studying the full micro-graph, of which Figure 2 was one subgraph, it turns out that one can decompose it into four subgraphs, i.e. ones having many internal but few external connections. This parsing can then be transferred to the global graph, as is indicated by the dashed lines in the figure. One conclusion from the figure is that not even the global structure is hierarchically ordered. That it is not can be seen by noting that the graph from left to right is more richly interconnected than would be a bottom-to-top hierarchy. On the other hand, the subgraphs are by no means unordered; the antecedence-consequence relation does impose a partial ordering. Except for the pair consisting of subgraphs 1 and 2, the ordering is complete, as can be seen from a table of connections among subgraphs:

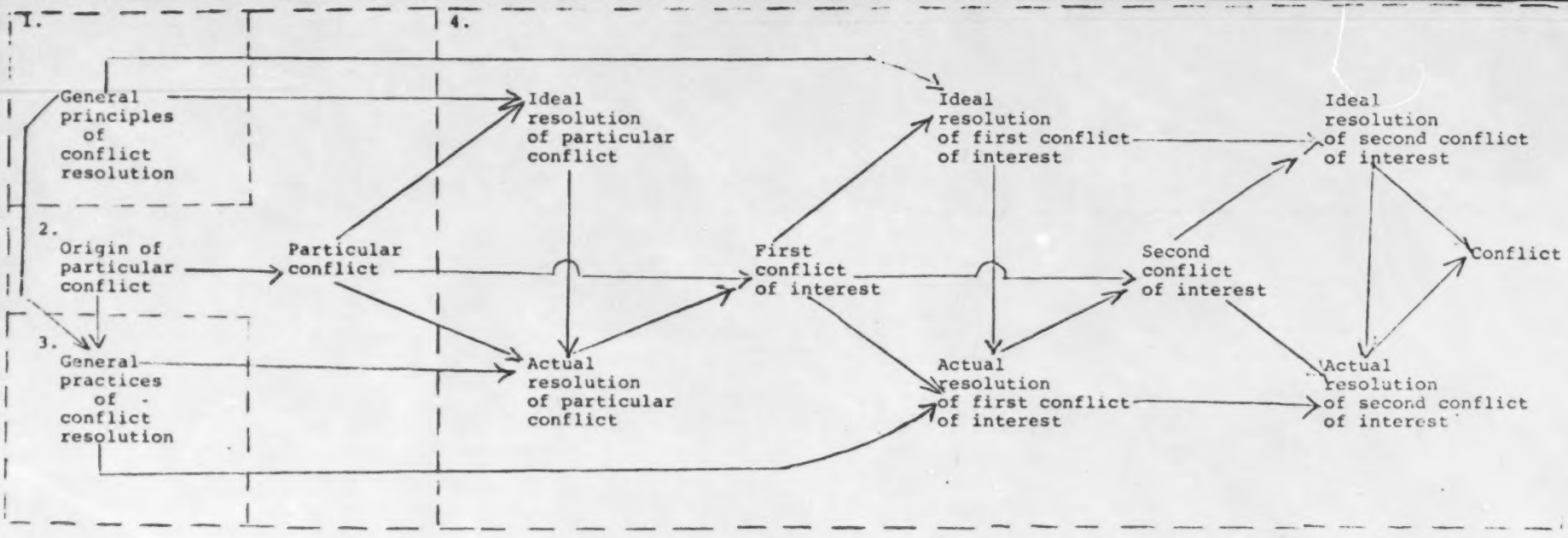



Figure 3. Global structure of Circle Island.

| | | Subgraph no. | | | |
|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 |
| Subgraph no. | 1 | X | - | X | X |
| | 2 | - | X | X | X |
| | 3 | - | - | X | X |
| | 4 | - | - | - | X |

Each "X" denotes a connection from row subgraph to column subgraph. Hence the partial order is <(1,2), 3,4>.

The above table reveals that this particular text's complexity, or at least the complexity of its connections among subgraphs, increases rather than remaining constant as the text develops. Instead of their being a gradual progression--for example, the $n^{th}$ subgraph receiving inputs from the $(n-1)^{st}$ and $(n-2)^{nd}$--a cumulative pattern is evinced. The first three subgraphs converge on the climatic fourth, where all of the conflicts come home to roost.

Now consider the four subgraphs one by one. Subgraph 1 represents the propositions on the government's general principles of conflict resolution. Part of the third paragraph of the story is devoted to this topic. In this label, "general" contrasts with "particular", the canal issue instantiation. Also, "principles" contrasts with "practices". That all senators were ranchers was not a constitutional requirement but instead part of the implementation.

Subgraph 2 is actually the largest one in terms of number of propositions, despite the fact that in Figure 3 its internal structure has not been shown in as much detail as that of subgraph 4. If this subgraph were to be characterized semantically rather than narratively, an apt name might be "geo-economic conditions", or something of the sort. Corresponding paragraphs of the quoted text are the first, second, fourth and fifth. Prominent in this subgraph are the parallel chains mentioned earlier. The subgraph ends on the arisal of the canal issue.

Subgraph 3 treats the habitual implementation of the government's principles. The reason for distinguishing the habitual from the constitutional is

that the two turn out to have quite different, even conflicting, implications for the canal issue instantiation next. This subgraph reflects the last part of the third paragraph, namely the senate's membership as a product jointly of the government's principles and the island's geo-economic conditions.

In subgraph 4, for sentences 22ff., are woven together the strands from the first three subgraphs. First the conflict between the farmers and ranchers is resolved; a vote is taken and a majority will decided. But this only leads to the senate's first conflict of interest, whose unsuccessful resolution sets off the second conflict of interest, which in turn leads to the potential civil war conflict.

In closing, it is worth reiterating that it was only after the micro-analysis, exemplified in the preceding figure, had been carried out that the overall structure (which scarcely resembles that described by published, more superficial analyses) could be inferred.

## VTII. Representing text interpretation by means of list graphs (Edward Crothers)

A new descriptive theory of text structure representation is recently being developed. Though at this early stage the approach appears highly promising, it would be premature to hazard any predictions about the eventual outcome of the undertaking. The theoretical direction is sufficiently different from current ones that considerable analysis is required to derive its first representations of texts, which can then form the basis of an initial estimate of the general merit of the method. These first representations, when completed, will be of the Aztecs and Pygmies passages already used in the experimental project on adjunct connectives.

The goal of the theory is a more powerful representation of text structure; that is, a representation which incorporates sources of textual cohesion normally recognized by a reader but not recognized in current approaches. (On the other hand, a satisfactory representation must not be too powerful relative to a reader's perception of a text. In other words, the theory must be flexible, able to accommodate text

structure ambiguities and differences of interpretation whenever readers individually or collectively do differ in their understandings of a text.) Naturally, a first step toward a more powerful theory is to demonstrate that such sources of inferable cohesion do exist and moreover do contribute importantly to the text's total understood cohesion. One, but not the only, prominent source of cohesion examined in this respect is parallelism. A graph description of parallelism in Circle Island has already been reported in the previous project. That work is now seen in retrospect as laying out the detailed, particular groundwork for the more abstract, general method of representation now in progress Basic to this abstraction is the notion of a list graph. Suffice it here to say that in this approach proposition lists, rather than individual propositions, are the nodes of the graph, and that a list can be expanded in several ways including abstraction. Another notable source of cohesion among propositions that the list graph description recognizes is the cohesion between a syntactically independent proposition and a syntactically subordinate recurrence of it. In particular, a frequent form of discourse development is the recounting or describing of an event "world", accompanied by subordinating it to another "world" which may be a world of desired events, known events, planned events, remembered events, etc. A special kind of relation is introduced in the theory to handle this subordinate, rather than parallel, source of cohesion. A final remark at this point is that the theory has implications for central issues of text structure, notably hierarchality and macrostructure, that have been addressed explicitly or implicitly by other theories.

References

Anderson, R. C. and Biddle, W. B. On asking people questions about what they are reading. In G. Bower (Ed.) Psychology of Learning and Motivation (Vol. 9). New York: Academic Press, 1975.

Corbett, A. T., & Wickelgren, W. A. Semantic memory retrieval: analysis by speed accuracy tradeoff functions. Quarterly Journal of Experimental Psychology, 1978, 30, 1-15.

Crothers, E. J. Inference and coherence. Discourse Processes, 1978, 1, 51-71.

Crothers, E. J. Text linguistics issues from a text representation perspective. In J. S. Petöfi (Ed.), Papers in Text Linguistics, Vol. 20. Text vs. sentence: basic questions of text linguistics, second part. Hamburg: Helmut, Buske, Verlag, 1979. (a)

Crothers, E. J. Paragraph structure inference. Norwood, N.J.: Ablex, 1979. (b)

Fetler, M. Using thematic information and text revisions to improve the effectiveness of adjunct questions. Unpublished doctoral dissertation, University of Colorado, Boulder, Colorado. 1978.

Frase, L. T. Prose processing. In G. Bower (Ed.) Psychology of Learning and Motivation (Vol. 9). New York: Academic Press, 1975.

Hicks, C. R. Fundamental concepts in the design of experiments (2nd ed.). New York: Holt, Rinehart, and Winston, 1973.

Johnson, R. E. Recall of prose as a function of the structural importance of the linguistic units. Journal of Verbal Learning and Verbal Behavior, 1970, 9, 12-20.

Kintsch, W. Comprehension and memory of text. In W. K. Estes (Ed.) Handbook of Learning and Cognitive Processes (Vol. 5). Hillsdale, N.J.: Erlbaum, 1977.

McClosky, M., & Glucksberg, S. Decision processes in verifying category membership statements: implications for models of semantic memory. Cognitive Psychology, 1979, 11, 1-37.

McKoon, G. Organization of information in text memory. Journal of Verbal Learning and Verbal Behavior, 1977, 16, 247-260.

Overall, J. E., & Spiegel, D. K. Concerning least squares analysis of experimental data. Psychological Bulletin, 1969, 72, 311-322.

Pedhazur, E. J. Coding subjects in repeated measures designs. Psychological Bulletin, 1977, 84, 298-305.

Rickards, J. P. Interaction of position and conceptual level of adjunct questions on immediate and delayed retention of text. Journal of Educational Psychology, 1976, 68, 210-217.

Rickards, J. P. and Di Vesta, F. J. Type and frequency of questions in processing textual material. Journal of Educational Psychology, 1974, 66, 354-362.

Rothkopf, E. Z. Structural text features and the control of processes in learning from written materials. In Freedle and Carroll (Eds.) Language Comprehension and the Acquisition of Knowledge. Washington, D.C.: V. H. Winston, 1972, 315-336.

Rothkopf, E. Z. Writing to teach and reading to learn: A perspective on the psychology of written instruction. In N. L. Gage (Ed.) The Seventy-fifth Yearbook of the National Society for the Study of Education. Chicago, Illinois: University of Chicago Press, 1976, 91-129.